

# SEARCH REQUEST FORM

Requestor's  
Name: \_\_\_\_\_

Serial  
Number: \_\_\_\_\_

Date: \_\_\_\_\_ Phone: \_\_\_\_\_ Art Unit: \_\_\_\_\_

## Search Topic:

Please write a detailed statement of search topic. Describe specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples or relevant citations, authors, keywords, etc., if known. For sequences, please attach a copy of the sequence. You may include a copy of the broadest and/or most relevant claim(s).

## STAFF USE ONLY

Date completed: 04-12-04  
Searcher: Belding C 2528  
Terminal time: 20  
Elapsed time: \_\_\_\_\_  
CPU time: \_\_\_\_\_  
Total time: 25  
Number of Searches: \_\_\_\_\_  
Number of Databases: 1

### Search Site

\_\_\_\_\_ STIC  
\_\_\_\_\_ CM-1  
\_\_\_\_\_ Pre-S

### Type of Search

\_\_\_\_\_ N.A. Sequence  
\_\_\_\_\_ A.A. Sequence  
\_\_\_\_\_ Structure  
\_\_\_\_\_ Bibliographic

### Vendors

\_\_\_\_\_ IG  
\_\_\_\_\_ STN  
\_\_\_\_\_ Dialog  
\_\_\_\_\_ APS  
\_\_\_\_\_ Geninfo  
\_\_\_\_\_ SDC  
\_\_\_\_\_ DARC/Questel  
☒ Other \_\_\_\_\_

This Page Blank (uspto)

STIC-Biotech/ChemLib

118960

From: Leffers, Gerald  
Sent: Wednesday, April 07, 2004 1:37 PM  
To: STIC-Biotech/ChemLib  
Subject: 09/430,590

Please interference search/search SEQ ID NO: 3 (~ 6.5 kb). Thank you. Gerry Leffers

*Gerald G. Leffers Jr., PhD*  
Primary Examiner, Art Unit 1636  
Remsen Building, Room 02A69  
(571) 272-0772

RECEIVED  
APR - 7 2004  
STIC

Searcher: \_\_\_\_\_  
Phone: \_\_\_\_\_  
Location: \_\_\_\_\_  
Date Picked Up: \_\_\_\_\_  
Date Completed: \_\_\_\_\_  
Searcher Prep/Review: \_\_\_\_\_  
Clerical: \_\_\_\_\_  
Online time: \_\_\_\_\_

TYPE OF SEARCH:

NA Sequences: \_\_\_\_\_  
AA Sequences: \_\_\_\_\_  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST (where applic.)

STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 11, 2004, 08:06:01 ; Search time 15972 Seconds

(without alignments)  
17438.163 Million cell updates/sec

Title: US-09-430-590E-3

Perfect score: 6426

Sequence: 1 tcttggtcttgacactatc.....agaagttatattccatca 6426

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenBank1:  
1: gb\_ba:\*  
2: gb\_hcg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_pl:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_strs:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_scs:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_hcg\_hum:\*  
31: em\_hcg\_in:\*  
32: em\_hcg\_oth:\*  
33: em\_hcg\_mus:\*  
34: em\_hcg\_pln:\*  
35: em\_hcg\_rtd:\*  
36: em\_hcg\_mam:\*  
37: em\_hcg\_yrt:\*  
38: em\_sy:\*  
39: em\_hgo\_hum:\*  
40: em\_hgo\_mus:\*  
41: em\_hgo\_oth:\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6426	100.0	6426	8 AF007776	AF007776 Candida a
2	6421.2	99.9	6980	8 AF050215	AF050215 Candida a
3	432	6.7	4872	8 CAL251464	AJ251464 Candida a
4	400	6.2	408	8 AF030556	AF030556 Candida a
5	363.8	5.7	5928	8 DHA439551	AJ439551 Debaryomy
6	339.2	5.3	1470	8 AF078809	AF078809 Candida a
7	166	2.6	1888	6 AX073190	AX073190 Sequence
8	107	1.7	110000	2 PFMA113_07	Continuation (8 of
9	105.6	1.6	128101	2 AC140550	AC140550 Medicago
10	105.6	1.6	146275	2 AL935272	AL935272 Danio rer
11	105.4	1.6	72880	8 AP006375	AP006375 Lotus cor
12	105.4	1.6	127731	8 AP006094	AP006094 Lotus cor
13	103.8	1.6	105887	8 AP006381	AP006381 Lotus cor
14	103.4	1.6	129571	2 AC137081	AC137081 Medicago
15	102.2	1.6	13434	2 AC135386	AC135386 Medicago
16	101	1.6	82726	2 AC137509	AC137509 Medicago
17	101	1.6	91075	2 AC140772	AC140772 Medicago
18	101	1.6	106730	8 AC144724	AC144724 Medicago
19	101	1.6	127857	2 AC133780	AC133780 Medicago
20	100.4	1.6	116807	2 BX890614	BX890614 Danio rer
21	99.8	1.5	105246	8 AC079131	AC079131 Arabidops
22	99.8	1.5	103637	8 AC079604	AC079604 Arabidops
23	99.4	1.5	80561	8 AC073555	AC073555 Arabidops
24	99.4	1.5	115995	8 AC146566	AC146566 Medicago
25	98.8	1.5	109155	8 ATT20K12	AL173898 Arabidops
26	98.8	1.5	133299	8 OSU000098	AL606586 Oryza sat
27	98	1.5	85961	8 AC079131	AC079131 Arabidops
28	98	1.5	103637	8 AC079604	AC079604 Arabidops
29	97.4	1.5	133637	2 AC147008	AC147008 Medicago
30	97.4	1.5	137850	2 AC133863	AC133863 Medicago
31	97.2	1.5	97146	8 AC007259	AC007259 Arabidops
32	96.4	1.5	4793	8 ATLECRCK	U08010 A.thaliana
33	96.4	1.5	13558	8 ATT93161	Y93161 Arabidops
34	96.4	1.5	91851	8 ATT6124	AL138659 Arabidops
35	96.4	1.5	110000	2 PFMA113_24	Continuation (25 of
36	96.4	1.5	22581	2 BX537105	BX537105 Danio rer
37	96	1.5	114434	8 AC147011	AC147011 Medicago
38	95.6	1.5	106411	8 AP004956	AP004956 Lotus cor
39	95.2	1.5	198743	2 BX530070	BX530070 Danio rer
40	95	1.5	151344	2 AC131343	AC131343 Oryza sat
41	94.6	1.5	82612	8 ATT3185	AL163491 Arabidops
42	93.8	1.5	70382	2 AC147471	AC147471 Medicago
43	93.8	1.5	105246	8 AC139747	AC139747 Medicago
44	93.8	1.5	146415	5 AL929469	AL929469 Zebrafish
45	93.8	1.5	193499	8 CINS08C94	AL732381 Oryza sat

#### ALIGNMENTS

RESULT 1  
LOCUS AF007776 6426 bp DNA linear PLN 21-NOV-1997  
DEFINITION Candida albicans retrotransposon pal, complete sequence.  
ACCESSION AF007776  
VERSION AF007776.1 GI:2636718  
KEYWORDS

#### SOURCE

ORGANISM Candida albicans

REFERENCE 1 (bases 1 to 6426)  
Matthews,G.D., Goodwin,T.J., Butler,M.I., Berryman,T.A. and  
Poulter,R.T.

pal, a highly unusual Ty1/copia retrotransposon from the

Pred. No. is the number of results predicted by chance to have a

pathogenic yeast *Candida albicans*  
 J. Bacteriol. 179 (22), 7118-7128 (1997)  
 MEDLINE  
 98037512  
 PUBMED  
 9371461  
 REFERENCE  
 2 (bases 1 to 6426)  
 Matthews, G.D., Goodwin, T.J.D., Butler, M.I., Berryman, T.A. and  
 Poulter, R.T.M.  
 TITLE  
 Direct Submission  
 JOURNAL  
 Submitted (11-JUN-1997) Department of Biochemistry, University of  
 Otago, Box 56, Dunedin, New Zealand  
 FEATURES  
 source  
 1..6426  
 /organism="Candida albicans"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:5476"  
 /transposon="retrotransposon pcal"  
 1..280  
 281..291  
 /note="minus-strand primer-binding site"  
 398..1372  
 /note="ORF1"  
 /codon\_start=1  
 /transl\_table=12  
 /product="GAG protein"  
 /protein\_id="AAC49879.1"  
 /db\_xref="GI:2636720"  
 /translation="MSSAKNDNEGKVMESVDOANATSKYDEHTKAPNNLFIKENDL  
 PKLAVNGKSDVKNMEKFKHYAVPYVLELDYNNKDKKPKKXVGGITFTGCLDM  
 CLOSTFDRFLIMISKPLQKKEANLIKAAVDATVTSKDYITTSKLSFVVEHLL  
 VVCYNLPYLSQVEKLEIEILYNTSNVDEYSLPNLIGVLYFNHVKESALSLPLN  
 IHASYSKMIQANDSTVLPSCSTIAEMCDHPYARLVIPSKRYELNLIIVLPAPR  
 KPKGPKENSESQCKSKSKRKNKKHPKNDKGEKEKEKEKTSSE"  
 1370..1372  
 /note="potentially suppressed stop codon"  
 <1373..6103  
 /note="ORF2"  
 /codon\_start=1  
 /transl\_table=12  
 /product="POL protein"  
 /protein\_id="AAC49878.1"  
 /db\_xref="GI:2636719"  
 /translation="KTGAASINCVMNHNCKTTPPYENSHSLNANINWNGKRN  
 KYLVDTGATISVNNKDIISNVKADILEVADGATLEADICGDLIRIGVISILIE  
 NTLVLPESFNLVSLKQIEERGFNVLITKESVIVNGVAPETIIASKNADADYMGQ  
 FSEBIECDPYDGLADMLSNADODKSKSMENSEYOEHDSSRALINSLEVDL  
 DVEISPYVEBOLPTGQNDIYNFHLMSNEMSIKILLQYCGVLHTEKESLOKJA  
 DCKVLLSNKQBSNHNHSEKSRHRHRLHCDLGPFSNNKWKVLTVIDEHTGYI  
 EGIITPKKVVLDLITQLKTNMNPKNVAFPRSDNAPEPPOPDLEPGLMRETTAA  
 YSPBLNGLAEVYKLIIDQITRIYVTLGPQLKIIYVVIQYSTIMINHTRRSLKGT  
 PYGCVQLSEGNFRFPALDCCVTFSALEKRYGVSTIGAFSSIMGVIGASQC  
 FSYVLLNNRCDIILSPNRLRSYEVINSYLNLSLTGMSNBYVPAEIGQASQA  
 QYEVGIVSESYNDNDVNMHPKESYVQASPTLTGMSNBYVNDPVOQTLEN  
 PDDFNPQLTSESHDWSEYKSDENPSPSLHETPGDNPVSKPQLTGTSVIGSKK  
 EPIITHTKIDAPSIQGRHRSSTAYGLSHQPTGPASESLSGTDFRGVYVKE  
 TVSEDMHSDPYPSARDEQNPESLANKNRYEKIDGNSISPBGDQVINSNV  
 EOSNVEITDAGNSPTQDEYSEGRILNQDITVTVAKVIENKISIPNSIDHTELA  
 TDSGNSNSTESDIQSKOEISPVINEKNTETIQHLESILADKLDFEYVNDLEIN  
 VINDIDIANPLPDENNDCQMESFNNHSMSPAKKYFPEKEVKEKATIGKRSQST  
 TDPRAIRIVANTGETKRIEPEKKEVPIITVNLKRSQKYSPEVSGGTIVNPKYSLA  
 VYNKIDYNDPGMISMAELKERSKDYEVPIPTGVRKISMWHTETIOLHDSVAYLN  
 RKSRCVHGNROKEKIDYDPSVSPVIDVLTITLTIIGELGIMTIOHLDVSAAYLN  
 ASITSNPIYVFPKSYVPLKKNHMLKRSYGLKSGLEKNSHNTIKRYVEDIPTQVL  
 HNDGPHIEBESYIYGLYVDILWGSQKYNIDPVDLRHEVKKFGEISNTL  
 GIEFRTESGYILQERFLKLLDFKLDDSGYKNIWIPNDKREKVALIREVNPEN  
 DFEKPNETLDPDAKKLYQSGVSLMAATNTPDISVYVNSIGSGSNPNVHDEK  
 LIYCLAYIKNSMGYHIEYKRNINIPKSFVIEGFSDFASFPGLDRKISGTLIYVNG  
 NLVQATKQOTVIAQSSAACMLANTMLKAIKIKHMLDGEVGIKICHQNDQV  
 IKVLNNYCHHPRPIDICYKFLROLINDKVSISIVYTNNDVADCMKCKLSRAKFAF  
 VEGMKRLDIEDNQSTIONATTAE"  
 1373..1380  
 /note="Region: purine-rich tract"  
 1381..1455  
 /note="Region: potential pseudoknot"

misc\_feature 3455..3465  
 /note="Region: polypurine tract 2"  
 misc\_feature 6136..6146  
 /note="Region: polypurine tract 1"  
 LTR 6147..6426  
 ORIGIN  
 Query Match 100.0%; Score 6426; DB 8; Length 6426;  
 Best Local Similarity 100.0%; Freq. No. 0;  
 Matches 6426; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 1 TGTGGTTGTGCACTATTTTGTGACAACTGATCAATGAATGATGTTATATGA 60  
 1 TGTGGTTGTGCACTATTTTGTGACAACTGATCAATGAATGATGTTATATGA 60  
 61 GAATGAAAATTTTTCATCAGCTCAGGATGATGAGAACTAATCTATTTGTGTG 120  
 61 GAATGAAAATTTTTCATCAGCTCAGGATGATGAGAACTAATCTATTTGTGTG 120  
 121 ATAAATGAAGGATGAATAATACCAATCCAGATATCAAGATAGAGAGAGAGATT 180  
 121 ATAAATGAAGGATGAATAATACCAATCCAGATATCAAGATAGAGAGAGAGATT 180  
 181 TCAATATATATCTTGTGAATAATACTTGTCTATATTCATATACACTGACGTGT 240  
 181 TCAATATATATCTTGTGAATAATACTTGTCTATATTCATATACACTGACGTGT 240  
 241 AAGCGTCAATCTCAGGATGAAGAAAGTTATATTCATCAGATTAAGAGTGTAT 300  
 241 AAGCGTCAATCTCAGGATGAAGAAAGTTATATTCATCAGATTAAGAGTGTAT 300  
 301 AATCATTTTGTGTCCTCAATTTAGCGTTGATTAATTCAGTCTCAGATTTGAT 360  
 301 AATCATTTTGTGTCCTCAATTTAGCGTTGATTAATTCAGTCTCAGATTTGAT 360  
 361 GATAGTTTGAAGTTTGAAGATGACGAATTTCAAGATGAGTCCGAAAGATGATGA 420  
 361 GATAGTTTGAAGTTTGAAGATGACGAATTTCAAGATGAGTCCGAAAGATGATGA 420  
 421 TAACGAAGGAAGGTCATGGAAGGTGATCAAGCTATATCTATTAAGTGTGATGA 480  
 421 TAACGAAGGAAGGTCATGGAAGGTGATCAAGCTATATCTATTAAGTGTGATGA 480  
 481 AATATCAAGGCTATGATCAATATGCTTTTCAATAATTAATGACTTAAGTTGGC 540  
 481 AATATCAAGGCTATGATCAATATGCTTTTCAATAATTAATGACTTAAGTTGGC 540  
 541 CCGCGGATTCAGAAAGCGTGATTAATGGAATGAATTAATTTCCAGCTTGC 600  
 541 CCGCGGATTCAGAAAGCGTGATTAATGGAATGAATTAATTTCCAGCTTGC 600  
 601 TTAACCCGATGTTTGAATTTTCTGACTATATCTTAAGATTAATTAAGTTAA 660  
 601 TTAACCCGATGTTTGAATTTTCTGACTATATCTTAAGATTAATTAAGTTAA 660  
 661 AAAGGATGAAGTATTTATTTTACGTTGTTGTTTCAATGTTTACAGTCAATTT 720  
 661 AAAGGATGAAGTATTTATTTTACGTTGTTGTTTCAATGTTTACAGTCAATTT 720  
 721 TGAATGTTCAATGATGATGATTTTACGTTGTTTCAATGTTTACAGTCAATTT 780  
 721 TGAATGTTCAATGATGATGATTTTACGTTGTTTCAATGTTTACAGTCAATTT 780  
 781 CTTAATCAAGTCTGTTATGATGCTGTTATCTTAATCTTAAGATTAATTAAGTTAA 840  
 781 CTTAATCAAGTCTGTTATGATGCTGTTATCTTAATCTTAAGATTAATTAAGTTAA 840  
 841 GATCTGCTGAAGTTTGAAGCTTGAAGATGATGATGATGATGATGATGATGATGAT 900  
 841 GATCTGCTGAAGTTTGAAGCTTGAAGATGATGATGATGATGATGATGATGATGAT 900  
 901 TTTGCTGAGTGTGAAGAACTTGAAGAACTTGAAGAACTTGAAGAACTTGAAGAACT 960  
 TTTGCTGAGTGTGAAGAACTTGAAGAACTTGAAGAACTTGAAGAACTTGAAGAACT

Db 901 TTGCTGCAAGGTGGAAGAGAACTTGAGGAAATCTACACACTTCAACGTTGTGCA 960  
 Qy 961 TGAGATATGCGTAGTCTTCCAAATCTCATAGTCAAGTCTTGTACTCAATCATGTGAA 1020  
 Db 961 TGAGATATGCGTAGTCTTCCAAATCTCATAGTCAAGTCTTGTACTCAATCATGTGAA 1020  
 Qy 1021 GAAATCAAGGCTTAAAGTTGTTTGTGAATATTCATGCTCATATCTCAAGTGAT 1080  
 Db 1021 GAAATCAAGGCTTAAAGTTGTTTGTGAATATTCATGCTCATATCTCAAGTGAT 1080  
 Qy 1081 TCAAGTCAACATGATACATCATGACTCCCAAGTGTCTCTACCAAGTGTGAAGATGTG 1140  
 Db 1081 TCAAGTCAACATGATACATCATGACTCCCAAGTGTCTCTACCAAGTGTGAAGATGTG 1140  
 Qy 1141 TGATCATCTGTGATATGCTAGATTGTTGACATTCACAGCAACAATRTGAATCTTAACT 1200  
 Db 1141 TGATCATCTGTGATATGCTAGATTGTTGACATTCACAGCAACAATRTGAATCTTAACT 1200  
 Qy 1201 TATGTTAGTTTACAGAGCACAGAGAAACCAAGAGGAAACAGAGGAACTCACTGGA 1260  
 Db 1201 TATGTTAGTTTACAGAGCACAGAGAAACCAAGAGGAAACAGAGGAACTCACTGGA 1260  
 Qy 1261 ACAATCTCAAAAAGAAAGAACTGAAATCAAGAAAGAAATAGAAACATCCAAATCAGA 1320  
 Db 1261 ACAATCTCAAAAAGAAAGAACTGAAATCAAGAAAGAAATAGAAACATCCAAATCAGA 1320  
 Qy 1321 TAAAGTAAAGGTGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAG 1380  
 Db 1321 TAAAGTAAAGGTGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAG 1380  
 Qy 1381 TGCTGCTTCTATTAATGTTGTAATGATATACATATATGACGAAACCAAGTTCAGT 1440  
 Db 1381 TGCTGCTTCTATTAATGTTGTAATGATATACATATATGACGAAACCAAGTTCAGT 1440  
 Qy 1441 AGAAAAATCTCATCTCTTAATGCTTCTTGAAGTAAAGTTTAAAGTTTAAAGTT 1500  
 Db 1441 AGAAAAATCTCATCTCTTAATGCTTCTTGAAGTAAAGTTTAAAGTTTAAAGTT 1500  
 Qy 1501 TAAAGTAAAGGTGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAG 1560  
 Db 1501 TAAAGTAAAGGTGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAG 1560  
 Qy 1561 ATTGCTGAATGTTAAGAGAGCAACATTAAGTTTCTGTTGCTGATGCTACATTAGA 1620  
 Db 1561 ATTGCTGAATGTTAAGAGAGCAACATTAAGTTTCTGTTGCTGATGCTACATTAGA 1620  
 Qy 1621 AGCAATATGTAATGTTGATATATATGAGAGTGTGCTGATGCTGATGCTGATG 1680  
 Db 1621 AGCAATATGTAATGTTGATATATATGAGAGTGTGCTGATGCTGATGCTGATG 1680  
 Qy 1681 TACATTTGATTTACAGAGAAAGTTCTTTAATCTTGTGAGTTGAAACAAATTTGAAGACG 1740  
 Db 1681 TACATTTGATTTACAGAGAAAGTTCTTTAATCTTGTGAGTTGAAACAAATTTGAAGACG 1740  
 Qy 1741 AGGATTTAATGTTCTTATTAATCAAGATCAAGTATTTATTAACCAAAATGTGCTCC 1800  
 Db 1741 AGGATTTAATGTTCTTATTAATCAAGATCAAGTATTTATTAACCAAAATGTGCTCC 1800  
 Qy 1801 TACATTTAATGTTCTTATTAATCAAGATCAAGTATTTATTAACCAAAATGTGCTCC 1860  
 Db 1801 TACATTTAATGTTCTTATTAATCAAGATCAAGTATTTATTAACCAAAATGTGCTCC 1860  
 Qy 1861 AGAATCTTTAGATGATATTTGATTAATGATGTTGAGATATGTTGCAATGTCTAA 1920  
 Db 1861 AGAATCTTTAGATGATATTTGATTAATGATGTTGAGATATGTTGCAATGTCTAA 1920  
 Qy 1921 CCAATATGCAAGATTAATCAAGTATTAATGATGTTGAGATATGTTGCAATGTCTAA 1980  
 Db 1921 CCAATATGCAAGATTAATCAAGTATTAATGATGTTGAGATATGTTGCAATGTCTAA 1980  
 Qy 1981 TAGTCTGAGCATTAAATTAATTTCTTTGACGAGGTTGATGTTTGAATTTTC 2040  
 Db 1981 TAGTCTGAGCATTAAATTAATTTCTTTGACGAGGTTGATGTTTGAATTTTC 2040

Qy 2041 CCCATATGAGATTTGAACAAATGCTACCAACTGAGATTAAGAACATATTTATTTTCCA 2100  
 Db 2041 CCCATATGAGATTTGAACAAATGCTACCAACTGAGATTAAGAACATATTTATTTTCCA 2100  
 Qy 2101 TTGATGTCAAATATGATATGATGATGATGATGATGATGATGATGATGATGATGAT 2160  
 Db 2101 TTGATGTCAAATATGATATGATGATGATGATGATGATGATGATGATGATGATGAT 2160  
 Qy 2161 CGTACTTCACTTCAAAAAGAGATCTTCAAAAAGATTTGCTGATGATGATGATGAT 2220  
 Db 2161 CGTACTTCACTTCAAAAAGAGATCTTCAAAAAGATTTGCTGATGATGATGATGAT 2220  
 Qy 2221 ATGGAATGCAACACAGAGATGACATATCATCTTCAAGAAAGAAAGCTCGAGAGACA 2280  
 Db 2221 ATGGAATGCAACACAGAGATGACATATCATCTTCAAGAAAGAAAGCTCGAGAGACA 2280  
 Qy 2281 TGAGAGACTTCAATGATGATCTCGTCCATTTAGGTCCGAAATTAACAGTGTATTT 2340  
 Db 2281 TGAGAGACTTCAATGATGATCTCGTCCATTTAGGTCCGAAATTAACAGTGTATTT 2340  
 Qy 2341 AACCTGTTATATGATGAAATATCGGTTACATTTGAAGATTTATTTAAAGACAGAA 2400  
 Db 2341 AACCTGTTATATGATGAAATATCGGTTACATTTGAAGATTTATTTAAAGACAGAA 2400  
 Qy 2401 GGTAAAGATCTCTTAAATCAAGATTAAGATTAAGATTAAGATTAAGATTAAGAT 2460  
 Db 2401 GGTAAAGATCTCTTAAATCAAGATTAAGATTAAGATTAAGATTAAGATTAAGAT 2460  
 Qy 2461 GGCATCTTCAAGAGATATGCTCTGAGTTCCACACCTTGTGATTTAGTGAAT 2520  
 Db 2461 GGCATCTTCAAGAGATATGCTCTGAGTTCCACACCTTGTGATTTAGTGAAT 2520  
 Qy 2521 CGGTATTTGAAGGAACTATGAGGAACTATGAGGAACTATGAGGAACTATGAGGAA 2580  
 Db 2521 CGGTATTTGAAGGAACTATGAGGAACTATGAGGAACTATGAGGAACTATGAGGAA 2580  
 Qy 2581 TGTATTAATATGATTTTCAACAGATTTACAGATGCTGTTGACACTTGTCCAAAT 2640  
 Db 2581 TGTATTAATATGATTTTCAACAGATTTACAGATGCTGTTGACACTTGTCCAAAT 2640  
 Qy 2641 ACTCAAGTATTTATTAATGATGATTAATGATTAATGATTAATGATTAATGATTA 2700  
 Db 2641 ACTCAAGTATTTATTAATGATGATTAATGATTAATGATTAATGATTAATGATTA 2700  
 Qy 2701 TCGTCACTCAAGGAGCAACCCCTTATGATGCTTATGATTAATGATTAATGATTA 2760  
 Db 2701 TCGTCACTCAAGGAGCAACCCCTTATGATGCTTATGATTAATGATTAATGATTA 2760  
 Qy 2761 CTACCGGTTTCTTTTGCATGATGATGATGATGATGATGATGATGATGATGATGAT 2820  
 Db 2761 CTACCGGTTTCTTTTGCATGATGATGATGATGATGATGATGATGATGATGATGAT 2820  
 Qy 2821 CGGTACGAGGATTAATCAATTAAGAGAGCTCTTATGATGATGATGATGATGAT 2880  
 Db 2821 CGGTACGAGGATTAATCAATTAAGAGAGCTCTTATGATGATGATGATGATGAT 2880  
 Qy 2881 CTACGATGAGATGATTTAGTATTAATGATGATGATGATGATGATGATGATGATGAT 2940  
 Db 2881 CTACGATGAGATGATTTAGTATTAATGATGATGATGATGATGATGATGATGATGAT 2940  
 Qy 2941 CTTAGCCTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3000  
 Db 2941 CTTAGCCTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3000  
 Qy 3001 CTTATCACTACACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3060  
 Db 3001 CTTATCACTACACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3060  
 Qy 3061 ACTGAGGCTCAGTACAGAGTACGAGGAAATATGAGAAAGTGAATATGACAAATACAA 3120  
 Db 3061 ACTGAGGCTCAGTACAGAGTACGAGGAAATATGAGAAAGTGAATATGACAAATACAA 3120

QY	3121	TGAGGTATGCAACAGCCCAAGAGTCAATTCAGTTAGCAGCATGTTTACTTTAAC	3180
Db	3121	TGACGTATGCAATGCCCCAAGAGTCAATTCAGTTAGCAGCATGTTTACTTTAAC	3180
QY	3181	TACGGGTACAGTTTCTAAAGAAATATGTTAATATGATATCCAGTACAGATTACATTTGA	3240
Db	3181	TACGGGTACAGTTTCTAAAGAAATATGTTAATATGATATCCAGTACAGATTACATTTGA	3240
QY	3241	GAATCCCATGATTTTTTCTAACCCCTTCAACTGAAGAAATCACAGATATGATATC	3300
Db	3241	GAATCCCATGATTTTTTCTAACCCCTTCAACTGAAGAAATCACAGATATGATATC	3300
QY	3301	CGAATGAATATCGATAGAAATCTTAAACCCAGTCTCCACAGCTTAAACACTGGGGATTA	3360
Db	3301	CGAATGAATATCGATAGAAATCTTAAACCCAGTCTCCACAGCTTAAACACTGGGGATTA	3360
QY	3361	TCCGCTGTCTAAACCTCTCAACTTGTATCCGAGCTTCAGTAAATAGGAAATCTTAAAG	3420
Db	3361	TCCGCTGTCTAAACCTCTCAACTTGTATCCGAGCTTCAGTAAATAGGAAATCTTAAAG	3420
QY	3421	GCCATTTCAAACCAACAGAGAGAGCCCTTCCATCCAGGGAGAGGACATTAACGCT	3480
Db	3421	GCCATTTCAAACCAACAGAGAGAGCCCTTCCATCCAGGGAGAGGACATTAACGCT	3480
QY	3481	GGAAATCTACTGCTCAGTTGGAATATCACACCACCCAGACTGGTACTCCCGCTTGCGA	3540
Db	3481	GGAAATCTACTGCTCAGTTGGAATATCACACCACCCAGACTGGTACTCCCGCTTGCGA	3540
QY	3541	GGAGTCAAAATGTCAGAGACAGATCATTTGGGTGAGACGTGTTAAAGAAACAGTCTC	3600
Db	3541	GGAGTCAAAATGTCAGAGACAGATCATTTGGGTGAGACGTGTTAAAGAAACAGTCTC	3600
QY	3601	AGAAATTTGGCATACTTCTGACTACCCAGAAATAGTCTGAGATGAAACAGCAAAATCC	3660
Db	3601	AGAAATTTGGCATACTTCTGACTACCCAGAAATAGTCTGAGATGAAACAGCAAAATCC	3660
QY	3661	CTCGTTACTGCTTAATAGAAATCCGGGTAACTGAAAATTAAGATGAGGAGAAATATTTCC	3720
Db	3661	CTCGTTACTGCTTAATAGAAATCCGGGTAACTGAAAATTAAGATGAGGAGAAATATTTCC	3720
QY	3721	ATTTCCGGGGGGTGTATGATGATTTCTGTGTATCACTCAATGTTTGAGCATCTAATGT	3780
Db	3721	ATTTCCGGGGGGTGTATGATGATTTCTGTGTATCACTCAATGTTTGAGCATCTAATGT	3780
QY	3781	TGAATCAGAGATGCTGGTACAGTCCAAATTCAGACGAAAGTTTCTCAAGGGAGAAAT	3840
Db	3781	TGAATCAGAGATGCTGGTACAGTCCAAATTCAGACGAAAGTTTCTCAAGGGAGAAAT	3840
QY	3841	AGTTATGACAAACCTGATATGTTGATACCTGTCCTAAAGTTATTTGAGATGAAAAAT	3900
Db	3841	AGTTATGACAAACCTGATATGTTGATACCTGTCCTAAAGTTATTTGAGATGAAAAAT	3900
QY	3901	CTCTCCTATTAAATTCATTAGATGATCATCTGAACTTGCTCAAGCTCCGGAAATGATAG	3960
Db	3901	CTCTCCTATTAAATTCATTAGATGATCATCTGAACTTGCTCAAGCTCCGGAAATGATAG	3960
QY	3961	CATTTCACAGATCCGATATTCATCGAAAAATATAATCACAGATGATTAATGAGAA	4020
Db	3961	CATTTCACAGATCCGATATTCATCGAAAAATATAATCACAGATGATTAATGAGAA	4020
QY	4021	AAATATGAAATATTCACAAAAACATTTGAAGTATCTTGTGCTGATTAAGATTTGATGA	4080
Db	4021	AAATATGAAATATTCACAAAAACATTTGAAGTATCTTGTGCTGATTAAGATTTGATGA	4080
QY	4081	ATTTCGAACGTATTAATGTTGATGAAATTTGAGAAATGATGATTAATGACATGCTGCTGA	4140
Db	4081	ATTTCGAACGTATTAATGTTGATGAAATTTGAGAAATGATGATTAATGACATGCTGCTGA	4140
QY	4141	AGCTATTCACATCCGATGAGAAATATATATGTTTCAAGATGAAATGAGATTTTGAATATA	4200
Db	4141	AGCTATTCACATCCGATGAGAAATATATATGTTTCAAGATGAAATGAGATTTTGAATATA	4200
QY	4201	TCAATGATGTACGAGCAAGAGAAATATCACTTTGAGAAAGATTTAACGAAAAAT	4260

Db	4201	TCATAGCATGTCACGACCAAGAAAGAAATACACATTGGAGAAAGAGTTACCGAAAAAT	4260
QY	4261	TGCTGGTACTAAACATTCACTTGATACAACTGATCCAAAGAAACAAATCAAGTGTTTAA	4320
Db	4261	TGCTGGTACTAAACATTCACTTGATACAACTGATCCAAAGAAAGCAATCAAGTGTTTAA	4320
QY	4321	TACTGGTGAACCAAGAGATCCGAACCCAAAGAAAGAGAGTGCCTTCACTGTGGAATT	4380
Db	4321	TACTGGTGAACCAAGAGATCCGAACCCAAAGAAAGAGAGTGCCTTCACTGTGGAATT	4380
QY	4381	AAACAAAAGATCCGAATACAAAGTACCATAATGTTACAAAGTGTGAAACGGTTATPAA	4440
Db	4381	AAACAAAAGATCCGAATACAAAGTACCATAATGTTACAAAGTGTGAAACGGTTATPAA	4440
QY	4441	CCCCAAGAGTATTACATGCGGTCCGTCACAAAATCGACTATATGATCCGGAGTGGAT	4500
Db	4441	CCCCAAGAGTATTACATGCGGTCCGTCACAAAATCGACTATATGATCCGGAGTGGAT	4500
QY	4501	AAAGTCATGATGCTGAACATAGAGAAATTTAGATCAAAAGATGTTTACGAGAGTTCC	4560
Db	4501	AAAGTCATGATGCTGAACATAGAGAAATTTAGATCAAAAGATGTTTACGAGAGTTCC	4560
QY	4561	AATTCACACCGGTGTGAAGCCTATATCTATGCGGTGCGTACATACAGAAAAATGATTC	4620
Db	4561	AATTCACACCGGTGTGAAGCCTATATCTATGCGGTGCGTACATACAGAAAAATGATTC	4620
QY	4621	TCTCAAGAGTGTGTGTGGAAATACAGGTGTGTGTGCATAGGCAACAGACAAAAGAAA	4680
Db	4621	TCTCAAGAGTGTGTGTGGAAATACAGGTGTGTGTGCATAGGCAACAGACAAAAGAAA	4680
QY	4681	ATTGGATTATGACCCCTTTTATGTGTTAAGTTCACCGTTATAGATCTTGACTATAGATT	4740
Db	4681	ATTGGATTATGACCCCTTTTATGTGTTAAGTTCACCGTTATAGATCTTGACTATAGATT	4740
QY	4741	ATTGACATAATAGTGTGTGAATTAAGAAATGACAATTCACATTTAAGCTGAGTGGC	4800
Db	4741	ATTGACATAATAGTGTGTGAATTAAGAAATGACAATTCACATTTAAGCTGAGTGGC	4800
QY	4801	GTATCTAAATGCCCTCTATTACTCAATTCATAATCCCAATTTATGTCTTCTCCTAATCAGT	4860
Db	4801	GTATCTAAATGCCCTCTATTACTCAATTCATAATCCCAATTTATGTCTTCTCCTAATCAGT	4860
QY	4861	ACCTTTGAAGAAAAACCATGTGTGTTATTGAAACGTTCTGTCTATGCGTTAAAAACATC	4920
Db	4861	ACCTTTGAAGAAAAACCATGTGTGTTATTGAAACGTTCTGTCTATGCGTTAAAAACATC	4920
QY	4921	GGGTTTGAAGATGATCACATATCAATCAAAAGAGATTGGAAGACATTTGTTTACTCAGT	4980
Db	4921	GGGTTTGAAGATGATCACATATCAATCAAAAGAGATTGGAAGACATTTGTTTACTCAGT	4980
QY	4981	TTTACACATGATGTTTATTTCACATTTGAATGAAGAGGATCAGTATATATTTAGG	5040
Db	4981	TTTACACATGATGTTTATTTCACATTTGAATGAAGAGGATCAGTATATATTTAGG	5040
QY	5041	TTTATATGTGATGATATCTTATAGTGTGAAGTCCAAAABTPTATGTAATTTTGT	5100
Db	5041	TTTATATGTGATGATATCTTATAGTGTGAAGTCCAAAABTPTATGTAATTTTGT	5100
QY	5101	GGATCAATTGAGATCATTTTGAAGTTAAAGTGTGTGTAATATCAATTAATCTTGG	5160
Db	5101	GGATCAATTGAGATCATTTTGAAGTTAAAGTGTGTGTAATATCAATTAATCTTGG	5160
QY	5161	TATTGAATTTGTTAAACCGAATCTGGTTATATTTTATCTCMAAAAAATTTCTCAAGAA	5220
Db	5161	TATTGAATTTGTTAAACCGAATCTGGTTATATTTTATCTCMAAAAAATTTCTCAAGAA	5220
QY	5221	ATTACTTAAGATTTCAAACTAGATGATCTATATGGGAAAAACATACCTTGATTTCCGAA	5280
Db	5221	ATTACTTAAGATTTCAAACTAGATGATCTATATGGGAAAAACATACCTTGATTTCCGAA	5280
QY	5281	TGACAAATATGAAAAAGTTGCAATPATTTCTGAAAAAGTTTATCCAGAGATGATTTTGA	5340



Db	5281	TGACAAATATGAAAGGTTGCATTAATTCGGAAGAAAGCTTATATCAGAGAAATGATTTTGA	5340
Qy	5341	AAAGSTTCGGAATGAGACATTCGTTGACCCCTGATGCTTAAAAAATTATACCAAAATGCTGT	5400
Db	5341	AAAGSTTCGGAATGAGACATTCGTTGACCCCTGATGCTTAAAAAATTATACCAAAATGCTGT	5400
Qy	5401	TGCGCTCCTTTATGAGGCTGCGACAAAACACAGCTCCAGATATATCGTCCGATGGAATTC	5460
Db	5401	TGCGCTCCTTTATGAGGCTGCGACAAAACAGCTCCAGATATATCGTCCGATGGAATTC	5460
Qy	5461	GTTGGGCTCTAAATCTGCAAAATCCAAATGTCATGATTTATGAGAAATTGATTTATTTGCT	5520
Db	5461	GTTGGGCTCTAAATCTGCAAAATCCAAATGTCATGATTTATGAGAAATTGATTTATTTGCT	5520
Qy	5521	TAGGATATCAAAATATAGCATGGGATATCATTCATGATGAGTAAAGAAACAGATTGAATAT	5580
Db	5521	TAGGATATCAAAATATAGCATGGGATATCATTCATGATGAGTAAAGAAACAGATTGAATAT	5580
Qy	5581	ACCAACCAAAATCATTTGTTTATCGAATGTTTCAGTATCGCTCATTTGCAACGAGATTGGA	5640
Db	5581	ACCAACCAAAATCATTTGTTTATCGAATGTTTCAGTATCGCTCATTTGCAACGAGATTGGA	5640
Qy	5641	TAGAAATCTATTTAGTGAACCTTTGATTTATGTGAATGGAAATTTGGTGCATGGGGAC	5700
Db	5641	TAGAAATCTATTTAGTGAACCTTTGATTTATGTGAATGGAAATTTGGTGCATGGGGAC	5700
Qy	5701	CAAAAAACAAACGCTATAGCACAAAGCTCAGACGTTGTGMAATGTTGGCTCTAAATTA	5760
Db	5701	CAAAAAACAAACGCTATAGCACAAAGCTCAGACGTTGTGMAATGTTGGCTCTAAATTA	5760
Qy	5761	TACAAATGTTGAAAGCTATCGCAATTAATAAAACCATTTAATGATTTGGGTTTGAAGTAG	5820
Db	5761	TACAAATGTTGAAAGCTATCGCAATTAATAAAACCATTTAATGATTTGGGTTTGAAGTAG	5820
Qy	5821	TAGATCATTTGTCATCAAGACAACCAAGCTGTGATTTAAAGTTTGAAGAAATTACTATTG	5880
Db	5821	TAGATCATTTGTCATCAAGACAACCAAGCTGTGATTTAAAGTTTGAAGAAATTACTATTG	5880
Qy	5881	TCACCACATCGACCAATAGATATCTGCTATAGATTTCTACGCCAATTTGATCAATGATAA	5940
Db	5881	TCACCACATCGACCAATAGATATCTGCTATAGATTTCTACGCCAATTTGATCAATGATAA	5940
Qy	5941	AGATTTTTCATATTCCTATGTGAGAGACAATGATTAATACCCGATGTATGATCAATAGT	6000
Db	5941	AGATTTTTCATATTCCTATGTGAGAGACAATGATTAATACCCGATGTATGATCAATAGT	6000
Qy	6001	TCTAAGTGTCTAAATCAAGACATTTGTTGAGGGTATGATTAACACGGTATGACCTAGA	6060
Db	6001	TCTAAGTGTCTAAATCAAGACATTTGTTGAGGGTATGATTAACACGGTATGACCTAGA	6060
Qy	6061	AGATATATCAAAACCTGATACAAATATGCAATTAACGCGAGAAATAGTGTATTTATCATTA	6120
Db	6061	AGATATATCAAAACCTGATACAAATATGCAATTAACGCGAGAAATAGTGTATTTATCATTA	6120
Qy	6121	ATTATCGTAATGCTCAATCAGGGGAGTGTGGTTTGTGCATATTTTGTGCAGAAACTG	6180
Db	6121	ATTATCGTAATGCTCAATCAGGGGAGTGTGGTTTGTGCATATTTTGTGCAGAAACTG	6180
Qy	6181	ATCAATGAAAAATGATGTTATTTATGAGATGGAAAAATTTTTCATCACACATCAGTGAT	6240
Db	6181	ATCAATGAAAAATGATGTTATTTATGAGATGGAAAAATTTTTCATCACACATCAGTGAT	6240
Qy	6241	GACGAACTAAACTATATTTGTGTGTATTAATTAAGGGTATGAATAATACCAATCCCGAGA	6300
Db	6241	GACGAACTAAACTATATTTGTGTGTATTAATTAAGGGTATGAATAATACCAATCCCGAGA	6300
Qy	6301	TATCAACGAGATAGAAAGGAGAGTTTCATATATATCTTGTGATTAATTAACCTTGCTCT	6360
Db	6301	TATCAACGAGATAGAAAGGAGAGTTTCATATATATCTTGTGATTAATTAACCTTGCTCT	6360
Qy	6361	AATTTACTATATACCAACTTGAACGTGTCAACGCTCAATCTCAGGTAAAGAAAGTTATATTT	6420
Db	6361	AATTTACTATATACCAACTTGAACGTGTCAACGCTCAATCTCAGGTAAAGAAAGTTATATTT	6420

```

RESULT 2
AF050215 6980 bp DNA linear PLN 01-JUN-1998
LOCUS      Candida albicans Tca2 retrotransposon gag polypeptide (gag) and pol
DEFINITION polypeptide (pol) genes, complete cds.
ACCESSION  AF050215
VERSION     AF050215.1  GI:3273716
SOURCE      Candida albicans
            Candida albicans
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1 (bases 1 to 6980)
            Goodwin, I.J.D. and Poulter, R.T.M.
            Temperature- and strain-dependent expression of the pCal
            retrotransposon of Candida albicans
            Unpublished
            2 (bases 1 to 6980)
            Goodwin, I.J.D. and Poulter, R.T.M.
            Direct Submission
            Submitted (20-FEB-1998) Department of Biochemistry, University of
            Otago, Cumberland Street, Dunedin, New Zealand
            Location/Qualifiers
FEATURES             source
                     1..6980
                        /organism="Candida albicans"
                        /mol_type="genomic DNA"
                        /strain="HOG759"
                        /db_xref="taxon:5476"
                        /chromosome="3"
                        /map="between CDRI and DYN1"
                        /note="5', target site duplication"
                        /note="15", target site duplication"
                        /rpt_family="Tca2 retrotransposon"
                        /rpt_type="dispersed"
                        /rpt_type="dispersed"
                        /note="5", long terminal repeat"
                        487..518
                        /note="Complements tRNA Arg (UCU) fragment"
                        604..1578
                        /gene="gag"
                        604..1578
                        /gene="gag"
                        /gene="gag"
                        /codon_start=1
                        /transl_table=12
                        /product="gag polypeptide"
                        /protein_id="AAC24820.1"
                        /db_xref="GI:3273717"
                        /translation="MSSAKDDNDGKWSYVDONAKISKYDEHIKAFNNLFIKENDL
                        PKAAVQKQSVDKNBEREKYFHAAYPVTEFLDYNKDKPKYKVGIGITGCLQDM
                        CUGIIFRFRILINISKPKHLOKRNATIKAYIDAYTSSKDYITISKISFVAVHEHL
                        VVCNPLDSQVEKEKEELIINTSNVDEYVSPNLIQGYLFNHYKSEBALSPFLN
                        IHAAYSKWIQADNDTVLPSCSSTIAEKCDDHDPYALVDIPSKYELNIVLSIPAPE
                        KFKPKPENSSSEQSKNSKRKNKGHPKSDNDKGEKEKEKTSE"
                        1576..1578
                        /gene="gag"
                        /note="potentially suppressed stop codon"
                        <1579..6309
                        /gene="pol"
                        <1579..6309
                        /gene="pol"
                        /note="includes protease, integrase, reverse
                        transcriptase, RNase H"
                        /codon_start=1
                        /transl_table=12
                        /product="pol polypeptide"
                        /protein_id="AAC24821.1"

```

```

/db_xref="GI:3273718"
/translation="KTGAASINCVMIHNCSTTPFVENSHTSNASLNVNFKGLRPN
KYLIVDTGATISVNNKIDLSNVADATIEVSVDAGALEDCIGDLIRGVGISTILE
NTLYIPSSFNVLVSLKQIERGFVLLITKESVIVFNQNVAPITBARKNADLYMPO
FSESLCEDFPYDGLADMLSNANDKDSNMENSYOCHDSSRALINSLREVLV
DVEISPYGEGLDPTGDKNDIVNFHLSNMSTIEKILGKYGGLVYHTEKSLCIA
DCKRTLSNKRSHNSHRKRSRHRHLCPTIGPFRNSKWTLYATSDHETCYI
EGITITKRYKVDLLIQRLKIMNRFNRYVPSDNPPEPSSDLAREIMETLAA
YSPELNGLAENVKLILQIYRIVYVGLPKLIVYVISTYIMNPRSLKQOT
PYGYVQLSENGFRFPFPAIDCVVTFESNALEKNRYGVTSKSPSSIMGVAYSADG
FSYVLLKNRCDILSPNVAILRSYVINSYLNLSITTMKSHIVPAEIOVROGA
QYEVGYVESEYDNTNDVMHMKESYVQPAFTLTGNSNEVYINDPVOITLEN
PDPSNPLQLTSESDNVSEKSDENKPSLHELPGDNVSKPGLGTSTVIGSX
EPITNHTKDAISIOGRDKHSESTAOVLSHOPOTGPASBESKLGSDHMGVDVNE
TVSEDMRTSDPYPTSAEDBOONPSLSNKRVTREKIDEGENISPRGGDDSVYVNE
EQSVNVEDDAGNSPIQDESOSGRLNEQDIDVTAKVLENKISPIFINSLDHETLA
VINDGNSNSTESDIQSNEISPVINEKNTBIIQKHIESIADRLKEFEPYNDLEIN
VINDDIAENAPLPDENNDVQMESFPNNHSMRARKCYFEKENVKLAGTSHLDT
TDPREAIRVNTGSTRKIERKQEVPIYVLLNKRQYKSYVYRSGVITNPKRYLVA
VNNKIDYNDPMGIMSNMALEKPSRKQVYEVPIPTGVKISMGWYTEKIDSLKGV
RKSRCVHGNRQKRLDYPDSVSPVIDLVTIRILITIGCELGMTIOHLDVSAVLN
ASTHNSPIYFPEKSVPLKKNHMLKRSVGLKOSGPEWHTTKRVLIEDIGENVL
HNDLPHI EYBESVITYGLVDDILNVSSQKYLIDFVQLDHFVKVKGISVYL
GIERKTESGILSIOEKFELKRLKDFELDSYGNLPIWINDYKEXVALIRENVEN
DPEKVPBETLIDPDAKLYOSGVSLMAANTRPDISVVNVLSGSKANPVHDXEK
LICYLRKSNMGYHIEYKANRLNIPKSEVIGCFSPASAPGLDRKSLSGTLIYNG
NLVOMATRKQTVIWOSSACEMALANTYMLKAEIKNMLDGLFEVKGKHCHODNCAV
IKULNNYCHHPRPIDI CYKELROLINDKXFSIYVNTNNVADMTKCLSRKXFAF
VEGKTRKLDIEDNOTSIQNAITAE"
misc_feature
1579..1586
/gene="pol"
/note="purine-rich tract"
misc_feature
1587..1661
/gene="pol"
/note="nucleos putative pseudoknot"
misc_feature
3661..3671
/gene="pol"
/note="polyurine tract 2"
misc_feature
6342..6352
/note="polyurine tract 1"
LTR
6353..6632
/note="3' long terminal repeat"
misc_feature
6633..6637
/note="3' target site duplication"

```

ORIGIN

Query Match 99.9%; Score 6421.2; DB 8; Length 6980;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 6423; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

QY 1 TGTGTTGTGCACTATTTTGTGTCAAACTGATCAATGAATAATGATGTTATTTGA 60
DB 207 TGTGTTGTGCACTATTTTGTGTCAAACTGATCAATGAATAATGATGTTATTTGA 266
QY 61 GAATGAAAAATTTTCCATCACAATCAGTGATGACAGAACTAACTATTTGTAGT 120
DB 267 GAATGAAAAATTTTCCATCACAATCAGTGATGACAGAACTAACTATTTGTAGT 326
QY 121 ATAAATTAAGGATGATAAATACCAATCCAGAAATATCAAGAGATGAAGAGAGATT 180
DB 327 ATAAATTAAGGATGATAAATACCAATCCAGAAATATCAAGAGATGAAGAGAGATT 386
QY 181 TCAATATATCTCTGTGTAATAATACTCGTCTCAATCACTATACCAACTAGAGGT 240
DB 387 TCAATATATCTCTGTGTAATAATACTCGTCTCAATCACTATACCAACTAGAGGT 446
QY 241 ACACGCTCAATCTCAGTAAAGAAAGTTATATTCATCAGATTGAAGTGCATAGTAT 300
DB 447 ACACGCTCAATCTCAGTAAAGAAAGTTATATTCATCAGATTGAAGTGCATAGTAT 506
QY 301 AATCATTTCGTCCTCAAAATTAAGCTTTGTAATAATCAGTCTCAGATTGTTATTTGAT 360
DB 507 AATCATTTCGTCCTCAAAATTAAGCTTTGTAATAATCAGTCTCAGATTGTTATTTGAT 566

```

```

QY 361 GATAGTTTGAAGTTGAAGGTACAGAAATTTCAAGATGAGTCCGCAAGAAATGATGA 420
DB 567 GATAGTTTGAAGTTGAAGGTACAGAAATTTCAAGATGAGTCCGCAAGAAATGATGA 626
QY 421 TAACCAAGGAAGGTCAAGGAAGTGTGATCAAGCTAATGCTATAGTAAGTGAAGA 480
DB 627 TAACCAAGGAAGGTCAAGGAAGTGTGATCAAGCTAATGCTATAGTAAGTGAAGA 686
QY 481 AATATCAAGGTGATATCAATATGCTTTTCAATAAATTTAATGACTTAAGTGGC 540
DB 687 AATATCAAGGTGATATCAATATGCTTTTCAATAAATTTAATGACTTAAGTGGC 746
QY 541 CGTCGATATCAAGAAACCGTGATTAATGAAATGAAGAAATTTAATTTCCAGTTC 600
DB 747 CGTCGATATCAAGAAACCGTGATTAATGAAATGAAGAAATTTAATTTCCAGTTC 806
QY 601 TTACCCGATGTTTGGAAATTTTGGCTTGACTATATCTTAAAGATAATTCAGGTTAA 660
DB 807 TTACCCGATGTTTGGAAATTTTGGCTTGACTATATCTTAAAGATAATTCAGGTTAA 866
QY 661 AAAGGTAGAGGTATTTATTTTAACTGTTGTTTCAATGATGTTTACAGTCCATTT 720
DB 867 AAAGGTAGAGGTATTTATTTTAACTGTTGTTTCAATGATGTTTACAGTCCATTT 926
QY 721 TGATAGTTTCAAGTATGATCATGATTTCTAAGCTACCAAGCACTTGCAAAAGCAAA 780
DB 927 TGATAGTTTCAAGTATGATCATGATTTCTAAGCTACCAAGCACTTGCAAAAGCAAA 986
QY 781 CTTATTCAAAGTCTTATGATGCTGTTACCAATCTTAATATATATCCATTAATGATGA 840
DB 987 CTTATTCAAAGTCTTATGATGCTGTTACCAATCTTAATATATATCCATTAATGATGA 1046
QY 841 GATCTTGTGAAGTTTGTAAAGCTTGAACATGAGTGTAGTGTGCTATTAACCTTCATA 900
DB 1047 GATCTTGTGAAGTTTGTAAAGCTTGAACATGAGTGTAGTGTGCTATTAACCTTCATA 1106
QY 901 TTGCTGCAAGTGAAGAAACCTTGAGAAATCTCTACAACTTCAACGTTGTCGA 960
DB 1107 TTGCTGCAAGTGAAGAAACCTTGAGAAATCTCTACAACTTCAACGTTGTCGA 1166
QY 961 TGAGTATGCTGATCTTCCAAATCTCATAGTCAAGTCTTGTGATCTTCAATCATGTGAA 1020
DB 1167 TGAGTATGCTGATCTTCCAAATCTCATAGTCAAGTCTTGTGATCTTCAATCATGTGAA 1226
QY 1021 GAATTCAGAGGCTTTAAGTTGTTTGTGAATTTCAATGCTCTCATATCTCAAAAGTGTAT 1080
DB 1227 GAATTCAGAGGCTTTAAGTTGTTTGTGAATTTCAATGCTCTCATATCTCAAAAGTGTAT 1286
QY 1081 TCAAGCTGACATGATATCATCATGATCTCCAAAGTGTCTTCCATAGCTGAAGAAATGTG 1140
DB 1287 TCAAGCTGACATGATATCATCATGATCTCCAAAGTGTCTTCCATAGCTGAAGAAATGTG 1346
QY 1141 TGATCATCTGATTTATGCTAAGTTGTTGACATTTCCAAACCAATATGAACTTAATCT 1200
DB 1347 TGATCATCTGATTTATGCTAAGTTGTTGACATTTCCAAACCAATATGAACTTAATCT 1406
QY 1201 TATGTTGTTTACAGACACAGAAACCAAAAGGAAACCAAGGGAAGCTCATGCGA 1260
DB 1407 TATGTTGTTTACAGACACAGAAACCAAAAGGAAACCAAGGGAAGCTCATGCGA 1466
QY 1261 ACAATCTCAAAAGAAAGAACTGAATCAAGAAAGAAATTAAGAAACATCCAAAATGAGA 1320
DB 1467 ACAATCTCAAAAGAAAGAACTGAATCAAGAAAGAAATTAAGAAACATCCAAAATGAGA 1526
QY 1321 TAACGATAAAGGTGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGG 1380
DB 1527 TAACGATAAAGGTGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGG 1586
QY 1381 TGCTGCTCTAATTAATTTGTTGATGATATATCATATATGTCAGCAAAACCAAGTTCCAGT 1440
DB 1587 TGCTGCTCTAATTAATTTGTTGATGATATATCATATATGTCAGCAAAACCAAGTTCCAGT 1646

```

QY 1441 AGAATTCCTATCTCTTAATGCTCTTGAACGTAAATTTAAAGGTTAAGGTT 1500  
DB 1647 AGAATTCCTATCTCTTAATGCTCTTGAACGTAAATTTAAAGGTTAAGGTT 1706  
QY 1501 TAAACAGATCTAGTGTATGATACGTGCGCAATATCTGTGTGAACTAAAGATAT 1560  
DB 1707 TAAACAGATCTAGTGTATGATACGTGCGCAATATCTGTGTGAACTAAAGATAT 1766  
QY 1561 ATTGCTGATGTTAAGACGCAACATTAAGTTCTGTGCTGATGCTGATTAAGA 1620  
DB 1767 ATTGCTGATGTTAAGACGCAACATTAAGTTCTGTGCTGATGCTGATTAAGA 1826  
QY 1621 AGCAGATGATGTTGATCTTAATTAAGAGTCGGTAATGCTGATTAAGAGAA 1680  
DB 1827 AGCAGATGATGTTGATCTTAATTAAGAGTCGGTAATGCTGATTAAGAGAA 1886  
QY 1681 TACATTTGATTTACAGAAAGTTCTTAATCTGTGATTTGAAACAAATTTGAAGA 1740  
DB 1887 TACATTTGATTTACAGAAAGTTCTTAATCTGTGATTTGAAACAAATTTGAAGA 1946  
QY 1741 AGGATTTAATGTTCTTAATTAAGAAATCAGTGTATTTAAACCAAAATGTCCTC 1800  
DB 1947 AGGATTTAATGTTCTTAATTAAGAAATCAGTGTATTTAAACCAAAATGTCCTC 2006  
QY 1801 TACTATTTATGCTTGAAGAAAGATGCTGATCTTAATGAGGTCTCTCAATTCAGGA 1860  
DB 2007 TACTATTTATGCTTGAAGAAAGATGCTGATCTTAATGAGGTCTCTCAATTCAGGA 2066  
QY 1861 AGAATCTTAGAATGATTTGATTTGATTAAGATGTTGAGATATGTTGCAATGCTAA 1920  
DB 2067 AGAATCTTAGAATGATTTGATTTGATTAAGATGTTGAGATATGTTGCAATGCTAA 2126  
QY 1921 CCAAGATGACAAAGATTAATCAAGTATGATTAAGATGCAAGATTCAGAAATGATTA 1980  
DB 2127 CCAAGATGACAAAGATTAATCAAGTATGATTAAGATGCAAGATTCAGAAATGATTA 2186  
QY 1981 TAGTCTGAGCATTAATTAATTTCTTTGACGAGGTGATGTTTGAATTTTC 2040  
DB 2187 TAGTCTGAGCATTAATTAATTTCTTTGACGAGGTGATGTTTGAATTTTC 2246  
QY 2041 CCCATATGAGTTGAACATTTGCTAACCACTGAGATTAAGAAAGATTTAAATTTCCA 2100  
DB 2247 CCCATATGAGTTGAACATTTGCTAACCACTGAGATTAAGAAAGATTTAAATTTCCA 2306  
QY 2101 TTTGATGTCAAATCATATGTCATGAGAAATCTGTTTGAAGAAATTCAGAGGCT 2160  
DB 2307 TTTGATGTCAAATCATATGTCATGAGAAATCTGTTTGAAGAAATTCAGAGGCT 2366  
QY 2161 CGTACTTCACTCACTCAAAAGAGTCTTCAAAAGATGCTGATTTGATGATGCTAAT 2220  
DB 2367 CGTACTTCACTCACTCAAAAGAGTCTTCAAAAGATGCTGATTTGATGATGCTAAT 2426  
QY 2221 ATGCAATGCAAAACGAGAAATGCAATCATCATTTCAAGAAAGAAAGCTCGAAGACA 2280  
DB 2427 ATGCAATGCAAAACGAGAAATGCAATCATCATTTCAAGAAAGAAAGCTCGAAGACA 2486  
QY 2281 TAGAGACTTCACTGATGATCTCTGCTGATTTGATGCTGAAATTAAGAGTGTATTT 2340  
DB 2487 TAGAGACTTCACTGATGATCTCTGCTGATTTGATGCTGAAATTAAGAGTGTATTT 2546  
QY 2341 AAGCTGTGTTATGATGAACATACGGGTTACATTGAAGAAATTTATTAATAAGCAGAAA 2400  
DB 2547 AAGCTGTGTTATGATGAACATACGGGTTACATTGAAGAAATTTATTAATAAGCAGAAA 2606  
QY 2401 GGTAAAGGATCTCTTAATCAACGATTAAGATCTGAATATCGGTTAAAGATTAAGT 2460  
DB 2607 GGTAAAGGATCTCTTAATCAACGATTAAGATCTGAATATCGGTTAAAGATTAAGT 2666  
QY 2461 GGCATATCTCAGAAAGTAAATGCTCTGAGTTCCTCAACACCTTCTGATTTAGCTGAGTT 2520  
DB 2667 GGCATATCTCAGAAAGTAAATGCTCTGAGTTCCTCAACACCTTCTGATTTAGCTGAGTT 2726  
QY 2521 CGGTATTTGAGAGAGATTAAGCGGCAATCTGCTGAGTTATGATGCTCGCGAGGT 2580

DB 2727 CGGTATTTGAGAGAGATTAAGCGGCAATCTGCTGAGTTATGATGCTCGCGAGGT 2786  
QY 2581 TGTAAATTAATGATTTTCAACAGATTTACAGAGATCTGTTGACACTGTTGCCAATAT 2640  
DB 2787 TGTAAATTAATGATTTTCAACAGATTTACAGAGATCTGTTGACACTGTTGCCAATAT 2846  
QY 2641 ACTCAAGTATTAATTAATGATTAAGTCAATATTTCAATTAAGATCAACCACTTCAG 2700  
DB 2847 ACTCAAGTATTAATTAATGATTAAGTCAATATTTCAATTAAGATCAACCACTTCAG 2906  
QY 2701 TCGTTACCTCAAGGCAAAACCCCTTAAGTGTCTATTAATTAAGAGGAAATTT 2760  
DB 2907 TCGTTACCTCAAGGCAAAACCCCTTAAGTGTCTATTAATTAAGAGGAAATTT 2966  
QY 2761 CTACCGGTTCTCTTTGCAATGATGTCGTTACATTTAGTAATGCCATGAAAAGAA 2820  
DB 2967 CTACCGGTTCTCTTTGCAATGATGTCGTTACATTTAGTAATGCCATGAAAAGAA 3026  
QY 2821 CCGTTACGAGATTAATCAATCAATTAAGAGCTCTTATGATCAAGGTCGTGATTTGG 2880  
DB 3027 CCGTTACGAGATTAATCAATCAATTAAGAGCTCTTATGATCAAGGTCGTGATTTGG 3086  
QY 2881 CTACGCTAGGATTTGTTAGTTATGATGCTGTTGCTAAATAATGCGGTGATTTAT 2940  
DB 3087 CTACGCTAGGATTTGTTAGTTATGATGCTGTTGCTAAATAATGCGGTGATTTAT 3146  
QY 2941 CTTTACCCCTAATGTCGTAATTTGCGAAGCTATGAGTTTAACTCTATCTCAAAA 3000  
DB 3147 CTTTACCCCTAATGTCGTAATTTGCGAAGCTATGAGTTTAACTCTATCTCAAAA 3206  
QY 3001 CTTATCACTACACCTATGTCACACATGTCCTAAGGTCGTAAGATCAAGGAAAGGA 3060  
DB 3207 CTTATCACTACACCTATGTCACACATGTCCTAAGGTCGTAAGATCAAGGAAAGGA 3266  
QY 3061 ACTGCGGCTCACTACAGAGTACGCGGAAATATGTAAGAAATATGAACATACAA 3120  
DB 3267 ACTGCGGCTCACTACAGAGTACGCGGAAATATGTAAGAAATATGAACATACAA 3326  
QY 3121 TGAAGTATGACATGCGCAAGATGATTAATGATGATGATGATGATGATGATGATGAT 3180  
DB 3327 TGAAGTATGACATGCGCAAGATGATTAATGATGATGATGATGATGATGATGATGAT 3386  
QY 3181 TACGGGTAACAGTTCTTAAGATGATTAATGATGATGATGATGATGATGATGATGAT 3240  
DB 3387 TACGGGTAACAGTTCTTAAGATGATTAATGATGATGATGATGATGATGATGATGAT 3446  
QY 3241 GAATCCGATGATTTTCTTAACCTCTCTTAACCTAATGATTAAGATTAAGATGATGAT 3300  
DB 3447 GAATCCGATGATTTTCTTAACCTCTCTTAACCTAATGATTAAGATTAAGATGATGAT 3506  
QY 3301 CGAAGTAAATGAGATGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 3360  
DB 3507 CGAAGTAAATGAGATGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 3566  
QY 3361 TCCGATGCTAAACCTCTCTCAACTTGTACGAGATGATGATGATGATGATGATGATGAT 3420  
DB 3567 TCCGATGCTAAACCTCTCTCAACTTGTACGAGATGATGATGATGATGATGATGATGAT 3626  
QY 3421 GCGTATTTCAACCAACCAAGAGAGCGCCCTTCACTCAAGGAGAGGACATTAAGCGCT 3480  
DB 3627 GCGTATTTCAACCAACCAAGAGAGCGCCCTTCACTCAAGGAGAGGACATTAAGCGCT 3686  
QY 3481 GGAATCTACTGCTCAGGTTGATCACTATCAACCAACCCAGATGATGATGATGATGATGAT 3540  
DB 3687 GGAATCTACTGCTCAGGTTGATCACTATCAACCAACCCAGATGATGATGATGATGATGAT 3746  
QY 3541 GGAATCTACTGCTCAGGTTGATCACTATCAACCAACCCAGATGATGATGATGATGATGAT 3600  
DB 3747 GGAATCTACTGCTCAGGTTGATCACTATCAACCAACCCAGATGATGATGATGATGATGAT 3806  
QY 3601 AGAAGATGAGCATCTTCTGATCAACCAAGAACTAATGCTGAAGATGAACAGCAAAATCC 3660

3897 AGAAGATTGGCACTACTTGTGACTACCCAGAACTAGTGTGAAGATGAACAGCAAAATCC 3866  
3661 CTCTTACTGTGCTAATAAGATCGGCTTACTGAAAATAATAGATGAGGAGAAAATATTTC 3720  
3867 CTCGTACTGGCTAATAAGATCGGCTTACTGAAAATAATAGATGAGGAGAAAATATTTC 3926  
3721 ATTTCCGGGGGGTGTATGATCTGTGTGATCACTCAATGTTGAGCATCTAATGT 3780  
3927 ATTTCCGGGGGGTGTATGATCTGTGTGATCACTCAATGTTGAGCATCTAATGT 3986  
3781 TGAACAGAGATGCTGTGAACAGTCCAAATCAAGCAAGATTCTCAAGAGGAGAAAT 3840  
3987 TGAACAGAGATGCTGTGAACAGTCCAAATCAAGCAAGATTCTCAAGAGGAGAAAT 4046  
3841 ACTTAATGAACAACCTGATATAGTGTACTGTGTAAGTTTGAATGAGATGAAAAT 3900  
4047 ACTTAATGAACAACCTGATATAGTGTACTGTGTAAGTTTGAATGAGATGAAAAT 4106  
3901 CTCTCTTATTAATCATTAGATGATCATCTGAACCTTGCTACAGACTCGGGAATGTAG 3960  
4107 CTCTCTTATTAATCATTAGATGATCATCTGAACCTTGCTACAGACTCGGGAATGTAG 4166  
3961 CAATTGCAACAAATCCGATTCATCGAAAATAAGAAATACACAGATTTAATGAGAA 4020  
4167 CAATTGCAACAAATCCGATTCATCGAAAATAAGAAATACACAGATTTAATGAGAA 4226  
4021 AAATACTGAATAATCCAAAACACATTTGAAGTATCTGTGATAGAGATGTGATGA 4080  
4227 AAATACTGAATAATCCAAAACACATTTGAAGTATCTGTGATAGAGATGTGATGA 4286  
4081 ATTGGAACGTAATATGTTGATGAAATTGAGAAATGTGATTAATGACATGACATGCTGA 4140  
4287 ATTGGAACGTAATATGTTGATGAAATTGAGAAATGTGATTAATGACATGACATGCTGA 4346  
4141 AGCTAATCCACTACACAGATGAATAATGATGTCAGATGATGAGAGTTTGTATTA 4200  
4347 AGCTAATCCACTACACAGATGAATAATGATGTCAGATGATGAGAGTTTGTATTA 4406  
4201 TCATAGCATGTACAGAGCAAGAAAGAAATACACATTTGAGAAAGATTAACGAAAAT 4260  
4407 TCATAGCATGTACAGAGCAAGAAAGAAATACACATTTGAGAAAGATTAACGAAAAT 4466  
4261 TGCTGTACTAATCACTTCACTGATCACTGATCCAAAGAAACATCCAGAGTTTAA 4320  
4467 TGCTGTACTAATCACTTCACTGATCACTGATCCAAAGAAACATCCAGAGTTTAA 4526  
4321 TACTGTGAAAACCAAGAAATCGAACCCAGAAAAGAGGTGCTTCACTGTGAAT 4380  
4527 TACTGTGAAAACCAAGAAATCGAACCCAGAAAAGAGGTGCTTCACTGTGAAT 4586  
4381 AAACCAAAAGATCGCAATACAGTCACTATGTTTCAAGAAAGTGTGAACGTTATTA 4440  
4587 AAACCAAAAGATCGCAATACAGTCACTATGTTTCAAGAAAGTGTGAACGTTATTA 4646  
4441 CCCCAAGAGTATTACATGCGGTGCTGCAACAAAATGACTATATGATCCGGATGAT 4500  
4647 CCCCAAGAGTATTACATGCGGTGCTGCAACAAAATGACTATATGATCCGGATGAT 4706  
4501 AAAGTCATGATGCTGAATCAAGAAATTTAGTCAAAAAGATTTTACGAAAGATTC 4560  
4707 AAAGTCATGATGCTGAATCAAGAAATTTAGTCAAAAAGATTTTACGAAAGATTC 4766  
4561 AATTCCACGAGTGAAGCTTATATCTATGAGTTGGTACATCTGAGAAAATGATTC 4620  
4767 AATTCCACGAGTGAAGCTTATATCTATGAGTTGGTACATCTGAGAAAATGATTC 4826  
4621 TCTCAAGGTGTGTTGCGAAATCAAGTTGTGTCATGCAACAGCAAAAGGAAA 4680  
4827 TCTCAAGGTGTGTTGCGAAATCAAGTTGTGTCATGCAACAGCAAAAGGAAA 4886  
4681 ATTGATTAATGACCTTTTATGATGATGATGATGATGATGATGATGATGATGATGAT 4740  
4887 ATTGATTAATGACCTTTTATGATGATGATGATGATGATGATGATGATGATGATGAT 4946

4741 ATTGAATTAATAGGTGTGGAATTAGAAATGCAATTCACATTTAGACGTGAGTGGC 4800  
4947 ATTGAATTAATAGGTGTGGAATTAGAAATGCAATTCACATTTAGACGTGAGTGGC 5006  
4801 GTATCTAAATGCTCTATTAATCAATTCATCAATTTATGTTCTCTCTAAATCAGT 4860  
5007 GTATCTAAATGCTCTATTAATCAATTCATCAATTTATGTTCTCTCTAAATCAGT 5066  
4861 ACCTTTGAAGAAAACCAATGTTGTTGTTTGAACGTTCTGATGAGGTTAAACAGTC 4920  
5067 ACCTTTGAAGAAAACCAATGTTGTTGTTTGAACGTTCTGATGAGGTTAAACAGTC 5126  
4921 GGGTTGGAATGTTATCACTATCAAAAGATTTGGAACATTTGTTTACTCAAGT 4980  
5127 GGGTTTGAATGTTATCACTATCAAAAGATTTGGAACATTTGTTTACTCAAGT 5186  
4981 TTTCACAAATGATGTTTATTTACATTTGATTAATGAAGAGGATCAGTAATATTTAG 5040  
5187 TTTCACAAATGATGTTTATTTACATTTGATTAATGAAGAGGATCAGTAATATTTAG 5246  
5041 TTTCATATGATGATGATTTATTTATGTTGAAAGTTCAAAAAGTTATTTATTTGT 5100  
5247 TTTCATATGATGATGATTTATTTATGTTTGAAGTTCAAAAAGTTATTTATTTGT 5306  
5101 GGATCAATGAGATCAATTTGAAAGTTTGAAGTTTGTGCAAAATTCATCTTGG 5160  
5307 GGATCAATGAGATCAATTTGAAAGTTTGAAGTTTGTGCAAAATTCATCTTGG 5366  
5161 TTTCATATTTGTTAAACCGAATCGTTATTTATTTATCTCAAGAAAATTTCTCAAGAA 5220  
5367 TTTCATATTTGTTAAACCGAATCGTTATTTATTTATCTCAAGAAAATTTCTCAAGAA 5426  
5221 ATTACTTAAGATTTCAAACTAGATGATGATGAGGAAAAACATPACCTGGAATCCGAA 5280  
5427 ATTACTTAAGATTTCAAACTAGATGATGATGAGGAAAAACATPACCTGGAATCCGAA 5486  
5281 TGACAAATTAATGAAGGTTTGCATTAATTCGAAAACGTTAATCCAGAAATGATTTTGA 5340  
5487 TGACAAATTAATGAAGGTTTGCATTAATTCGAAAACGTTAATCCAGAAATGATTTTGA 5546  
5341 AAAGGTTCCGATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5400  
5547 AAAGGTTCCGATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5606  
5401 TGCCCTGCTTTTATGAGGCTGCCCAACACACGTTCCAGATATATCGGTGAGTGAATTC 5460  
5607 TGCCCTGCTTTTATGAGGCTGCCCAACACACGTTCCAGATATATCGGTGAGTGAATTC 5666  
5461 GTTGGGTTCTAAATGTCAAAATCCAAATGTCATGATTAATGAAATTTGATTTGCT 5520  
5667 GTTGGGTTCTAAATGTCAAAATCCAAATGTCATGATTAATGAAATTTGATTTGCT 5726  
5521 TAGTATATCAAAAATAGCATGAGTATCACTTGAATCAAAAAGAAACAGATTGAATAT 5580  
5727 TAGTATATCAAAAATAGCATGAGTATCACTTGAATCAAAAAGAAACAGATTGAATAT 5786  
5581 ACCACCAAAATCATTTGTTATGCAATGTTTGAAGTATGCAATTTGACCAAGAGATGGA 5640  
5787 ACCACCAAAATCATTTGTTATGCAATGTTTGAAGTATGCAATTTGACCAAGAGATGGA 5846  
5641 TAGAAATCTATTAATGCAATTTGATTTATGTAATGTAATTTGTCATAGGCGAC 5700  
5847 TAGAAATCTATTAATGCAATTTGATTTATGTAATGTAATTTGTCATAGGCGAC 5906  
5701 CAAAAAACAACGGTCATAGCAACAAAGTCCAGCTTGTGAATTTGTTGGCTTAATTA 5760  
5907 CAAAAAACAACGGTCATAGCAACAAAGTCCAGCTTGTGAATTTGTTGGCTTAATTA 5966  
5761 TACATGTTGAAAGGATTCGAATTAATAAACAATTTAATGATTTGGTTTGAAGTAG 5820  
5967 TACATGTTGAAAGGATTCGAATTAATAAACAATTTAATGATTTGGTTTGAAGTAG 6026

Oy		5821	TAAAGTATCATTTGTCATCAAGAACAACCAAGCTGTGAATTAAAGTTTTGACAAAATAACTATCT	5889
Oy		6027	TAAAGTACATTTGTCATCAAGAACAACCAAGCTGTGAATTAAAGTTTTGACAAAATAACTATCT	6086
Oy		5881	TCACCACATCGACCAATAGATATATCTGCTATAGTTTCTAACGCCAATTGATCAATGATTA	5940
Db		6087	TCACCACATCGACCAATAGATATATCTGCTATAGTTTCTAACGCCAATTGATCAATGATTA	6146
Oy		5941	AGTATTTTGCATATCCCTATGTGAAGACAAATGATTAATAGCCGATTTGATGACTTAAGT	6000
Db		6147	AGTATTTTGCATATCCCTATGTGAAGACAAATGATTAATAGCCGATTTGATGACTTAAGT	6206
Oy		6001	TCTAAGTGTGCTAAATTCAAAGCATTGGTAGGGGTATGATTAACCGITTAGACTTGA	6060
Db		6207	TCTAAGTGTGCTAAATTCAAAGCATTGGTAGGGGTATGATTAACCGITTAGACTTGA	6266
Oy		6061	AGATPATTAACAACCTGATACAAATATGCAATAACGCGAGATAATAGTGGATTTATCATTA	6120
Db		6267	AGATPATTAACAACCTGATACAAATATGCAATAACGCGAGATAATAGTGGATTTATCATTA	6326
Oy		6121	ATTATCGTAATGCTCAATCAGGCGAGTGTGGTTGTGCACATATTTGTGCAGAACTG	6180
Db		6327	ATTATCGTAATGCTCAATCAGGCGAGTGTGGTTGTGCACATATTTGTGCAGAACTG	6386
Oy		6181	ATCAATGAAAAATGATGTTATTATGAGATGAAAAATTTTTCATCACACATCAGTGAT	6240
Db		6387	ATCAATGAAAAATGATGTTATTATGAGATGAAAAATTTTTCATCACACATCAGTGAT	6446
Oy		6241	GACGAACTTAACCTATTTGTGTGCTATTAATTAAGGGATTAATAATACCAATCCCCAGA	6300
Db		6447	GACGAACTTAACCTATTTGTGTGCTATTAATTAAGGGATTAATAATACCAATCCCCAGA	6506
Oy		6301	TATCAACGAGATAGAGAGGAGAGTTTCAATATATATCTTGATGAATAATAACTTGGTCT	6360
Db		6507	TATCAACGAGATAGAGAGGAGAGTTTCAATATATATCTTGATGAATAATAACTTGGTCT	6566
Oy		6361	AATTCATCTATACACAACCTTAGACGTGTACACGCTCAATCTCAGGTAAAGAACTTTATTT	6420
Db		6567	AATTCATCTATACACAACCTTAGACGTGTACACGCTCAATCTCAGGTAAAGAACTTTATTT	6626
Oy		6421	CCATCA 6426	
Db		6627	CCATCA 6632	
RESULT 3				
CAL251464/c				
LOCUS	CAL251464	4872 bp	DNA	linear
DEFINITION	Candida albicans eng1 gene for endo-1,3-beta-glucanase and pcal			PLN 08-JUN-2002
ACCESSION	Candidate retrotransposon.			
VERSION	AJ251464			
KEYWORDS	AJ251464.1 GI:6562333			
SOURCE	endo-1,3-beta-glucanase; eng1 gene; retrotransposon.			
ORGANISM	Candida albicans			
REFERENCE	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
AUTHORS	Saccharomycetales; mitosporic Saccharomycetales; Candida.			
TITLE	Esteban,P.F., García,R., Ríos,I., Vazquez De Aldana,C.R. and del Rey,F.			
REFERENCE	Cloning and characterization of the			
AUTHORS	endo-1,3-beta-glucanase-encoding gene CaENG1 in the yeast Candida			
TITLE	albicans			
JOURNAL	Unpublished			
REFERENCE	2 (bases 1 to 4872)			
AUTHORS	del Rey,F.			
TITLE	Direct Submission			
JOURNAL	Submitted (07-DEC-1999) Del Rey F., Departamento de Microbiología y			
FEATURES	Genética, Universidad de Salamanca, Campus Miguel de Unamuno,			
Source	Salamanca, 37007, SPAIN			
	Location/Qualifiers			
	1..4872			
	/organism="Candida albicans"			

[illegible]

Db 4452 TATATTCATCA 4441

RESULT 4  
LOCUS AF030556  
DEFINITION Candida albicans retrotransposon pCal LTR, linear PLN 01-JUL-1998  
ACCESSION AF030556  
VERSION AF030556.1 GI:3273502  
KEYWORDS  
SOURCE Candida albicans  
ORGANISM Candida albicans  
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
Saccharomycetales; mitosporic Saccharomycetales; Candida.

REFERENCE  
AUTHORS Goodwin,T.J.D. and Poulter,R.T.M.  
TITLE Temperature- and strain-dependent expression of the pCal retrotransposon of Candida albicans

JOURNAL  
REFERENCE Unpublished  
AUTHORS Goodwin,T.J.D. and Poulter,R.T.M.  
TITLE Direct Submission  
JOURNAL Submitted (21-OCT-1997) Department of Biochemistry, University of  
Otago, P.O. Box 56, Dunedin, New Zealand

FEATURES  
source  
1..408  
Location/Qualifiers  
/organism="Candida albicans"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:5476"  
/transposon="retrotransposon pCal"

LTR  
primer\_bind  
281..312  
/standard\_name="minus-strand primer binding site"  
/note="Perfect match to internal region of Candida albicans transfer RNA-Arg-UCU"

stem\_loop  
324..396  
398..5408  
/note="Start of GAG ORF"  
/codon\_start=1  
/transl\_table=12  
/product="GAG"  
/protein\_id="AAC24759.1"  
/db\_xref="GI:3273503"  
/translation="MSSA"

ORIGIN

Query Match 6.2%; Score 400; DB 8; Length 408;  
Best Local Similarity 98.8%; Pred. No. 1.1e-53;  
Matches 403; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 TGTGGTGTTCGCACTATTTGTGTCGAAACGATCAATGAAATGATGGTTATTTGA 60  
Db 1 TGTGGTGTTCGCACTATTTGTGTCGAAACGATCAATGAAATGATGGTTATTTGA 60  
QY 61 GAATGAAATTTTCCATCACATCAGTGTGATGACAGAACTAATATTTGTAGT 120  
Db 61 GAATGAAATTTTCCATCACATCAGTGTGATGACAGAACTAATATTTGTAGT 120  
QY 121 ATTAATTAAGGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 180  
Db 121 ATTAATTAAGGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 180  
QY 181 TCAATATATATCTTGTGAATATATTAATCTGTTCAATTCATATACAACTAGAGCTGT 240  
Db 181 TCAATATATATCTTGTGAATATATTAATCTGTTCAATTCATATACAACTAGAGCTGT 240  
QY 241 ACACCTCAATCTCAGGTAAAGAAAGTTATATTCATCAGATTGAAGTCCGATAGAT 300  
Db 241 ACACCTCAATCTCAGGTAAAGAAAGTTATATTCATCAGATTGAAGTCCGATAGAT 300  
QY 301 AATCATTTGTCGCCAAATTAAGGCTGTATTAATTAATTCAGTCCGAGATTGTATATGATT 360

Db 301 AACCAATTCGCCCAATTAAGGCTGTATTAATTAATTCAGTCCGAGATTGTATATGATT 360

QY 361 GATAGTTTCGAAGTTTGAAGTACAGATTTCAACAGATGAGTCCGC 408  
Db 361 GATAGTTTCGAAGTTTGAAGTACAGATTTCAACAGATGAGTCCGC 408

RESULT 5  
LOCUS DHA439551  
DEFINITION Debaryomyces hanseoni var. hanseoni LTR-retrotransposon Tdh2.  
ACCESSION AJ439551  
VERSION AJ439551.1 GI:20152518  
KEYWORDS gag gene; integrase; long terminal repeat; LTR; pol gene; protease;  
reverse transcriptase; RNaseH.  
SOURCE Debaryomyces hanseoni var. hanseoni  
ORGANISM Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
Saccharomycetales; Saccharomycetaceae; Debaryomyces.

REFERENCE  
AUTHORS Neuveglise,C., Feldmann,H., Bon,E., Galliardin,C. and Casaregola,S.  
TITLE Yeasts  
Genomic evolution of the LTR-retrotransposons in hemiascomycetous  
Yeasts

JOURNAL  
REFERENCE Unpublished  
AUTHORS Neuveglise,C.  
TITLE Direct Submission  
JOURNAL Submitted (18-MAR-2002) Neuveglise C., Genetique Molculaire et  
Cellulaire, Inra, Thiverval-Grignon, 78850, FRANCE

FEATURES  
source  
1..5928  
Location/Qualifiers  
/organism="Debaryomyces hanseoni var. hanseoni"  
/mol\_type="genomic DNA"  
/strain="CBS767"  
/variety="hanseoni"  
/db\_xref="taxon:58641"  
/transposon="LTR-retrotransposon Tdh2"

repeat\_region  
1..5928  
/transposon="LTR-retrotransposon Tdh2"

LTR  
primer\_bind  
360..394  
411..421  
/gene="gag"  
411..421  
/gene="gag"  
/codon\_start=1  
/transl\_table=12  
/protein\_id="CAD29536.1"  
/db\_xref="GI:20152518"  
/db\_xref="SPTREMBL:Q8TFK1"  
/translation="MSEQFTYSDPTVHVKHAKTVNNRYESSNEEMGSKTYKIHHSKS  
SDVINKPSSSLKFNQKEIVDNAIEQAKELVOCYFEKIPVHVDVSTLLILONAN  
KLSNKGIGKLLINDPTIDYGVKSSVYCONNNQIWPVPERELLSQSLFKHTS  
ALOKYPRVFSINVTYKYIRDNREVFSEILKLINRYIHAMNPFMDVTDISATAT  
VGTGEIPLINDPKLIDYRILILYRKLTKNSPAFIDELTPAVLIEPNNLL  
NDQILVNLKRTIRFKRYFOEKTDYNSNFDDNKKKQSKNNKYEHRNNSKXND  
EN"

gene  
1422..5444  
/gene="pol"  
/gene="pol"  
/note="contains protease, integrase, reverse transcriptase  
and RNaseH"  
/codon\_start=1  
/transl\_table=12  
/protein\_id="CAD29537.1"  
/db\_xref="GI:20152520"  
/db\_xref="GOA:O8TFK0"  
/db\_xref="SPTREMBL:Q8TFK0"  
/translation="QIGSHNTDIXSSFINNISDPPINNOYTYDNKKIDSKNY  
FDSGASVTVVDIKQILINIKEDLSESLFTVSNVPIYAAKEDLVYIKNKFTIN  
NAVYIPGIQSNLISFNDLNGFLILPNDYLPFKSNGSCFPIAKDNTSNLFNNI  
RSRDFSMKLSRKILIDELNNMOKIINKKTLNYPNNKYIKISNNHETIANPFKRD.

MYHLMGNMMSLESKXKYLKSGHIMKSESLTASBEERYKSCNEGLAINSKOSHHHT  
 FAPRPLRILHSDTLGIFSHRGRKYITITLIDEXGYANTTMSHKSLQDLPEKIRI  
 MNKRVADVNAFRTDNALDEMTKQDLIEKDEINSISPELNGISERNKSLIOP  
 IRKALLPQDRTLLPLKIVDYVYINMTPVRKSGSCPEALFESRNPISFIOF  
 GLDVLKSLSTEAKRYGVNSKSPCLFGFIIQYGDVHYKVLISTKXPPIIVP  
 NITLKSMNLKYLKSLDYLODKADIDISLGLDRTDEKSLIAEHNOPTIYE  
 CNEYDQESLVNSNSTIODVAOHIFERCHESNLSGMSPNVSNMTWTDNDYIT  
 EETNEVNSPDSPTDETTDSNITKTNEVNSDSNSDCKSSDTHPQASDKND  
 EINKNEGNTLSLESSTLSKPTPTQNNVRVRSKALIVETVERPKITSSKDR  
 VKDPDKTSLRNSKSKSLKNDQSHAKRVKSLFTIYVITVMSKSLNSQSKNS  
 SLVTRKINIVKRVKDLTDNMKQSIQRELDLTKYEVYVKNVNPVPIPTVWTH  
 KINLKEVQYKRCVVOGFRQJANEHYDTSKVSFVLSLILTLIAJAVEYIHL  
 LDISAVYADIDYEKSLFVKPPSGNSIDSGKQMLNKSVMQAGVWOCCTKYL  
 MDLNEPDTAISGMCKYGENKULIVLYVDMDLNSNTIINDPFLBELAKHFDK  
 YPADISEPLGIERIOLAGVYRISOHNFINSVYKKPLNNNGKXIPITKERNKSE  
 LKSEFTNBEVIRKESLNEBEDKELQSVGSLIMRANTRPDISFANQSLNSNNT  
 SVDLKLIYCLRYKQTLISFLEYRNRFSHKSGPILQTSDFSPAVTKDRSITGY  
 TVYLNGLINMSTIRQKXITDSSACEINALHSAVRSLKSKROAILDLNIVIDEITL  
 EDNAVIANCONNEGTSYGRWVDIKELRIQVSEGLIKLKYVNTSINIADMLTKALS  
 RCLFENRSLFERNDLKE"

5550..5928

ORIGIN  
 LTR

Query Match 5.7%; Score 363.8; DB 8; Length 5928;  
 Best Local Similarity 52.5%; Pred. No. 2.8e-48;  
 Matches 872; Conservative 0; Mismatches 777; Indels 12; Gaps 3;

4444 CAAGAGTATTACATGCGGCGTCAACAAATGACATATAATGATCCGGAGTAA 4503  
 3824 CACTGGAAAAATTAATACGTATATCGAAAGTCATTTAACAGATTAACAATGGAGCA 3883  
 4504 GTCAATGAATGCTGAAGTAAAGATTTAGATCAAAAGATTTTACGAAGAGTTCCAT 4563  
 3884 GTCAATACACGCGAATAGATCTTTAAGAAATATGAATATACGGTTGGAAAA 3943  
 4564 TCCCAACGGTGAAGCCATATCTATGAGTTGGTACATCTAGTGAATTTGATTTCT 4623  
 3944 TCTTAAAAATGCAACCTATTCACATCTGGGTTCTATACATCACTAAATTAACGACT 4003  
 4624 CAAGAGTGTGTCGGAATCACTGTTGTCATGCGCAACAGCAAAAGAAAAAT 4683  
 4004 CAAAGAGTTCAGTAAATACGTTGGTGTGAAGGCTTTAGCAAAATGGCAATGA 4063  
 4684 GGATTAATGACCTTTTATGTTAGTTACCTGTTATAGATCTTGACTATTAAGATT 4743  
 4064 ACACATATATACCTGGAAGGTCTATCTCTGTGATTAATATCCAAATTTCTTTACT 4123  
 4744 GACATTAATAGTGTGAATTAAGATGAATCAATTCATTTAGACGTCGAGTGGGCTA 4803  
 4124 TACAGCGATAGAGTTGAATTAAGATGCGCATCATCTTGATATATCTTCGCAATA 4183  
 4804 TCTAATGCTCTATTAATCAATCAATTTATGTTCTTCTCTTAATCAGTAACC 4863  
 4184 TTTACATGCAATATGACATAGAGAAATCCATATTTGTAAACACCACTGATGAA 4243  
 4864 TTTGAAGAAAAACCTGTGTGTTATTTGAAGCTTCTCTATGGTTAAACAGTCGGG 4923  
 4244 TATTATCTTGTGAATGTTGGCAATTAACAATCTTTTATGGAAGAAAAACAAGG 4303  
 4924 TTTGAATGATATCACTATCAAAAGATTTGGAAGACATGGTTTACTCAAGTTT 4983  
 4304 GATATATGATATCAATGATTAAGTATTAAGATCTTATTTTGAAGCTGATAC 4363  
 4984 ACACATATGATGTTTATTTACATGATATGAAGAGGATCAGTAATATTTAGGTT 5043  
 4364 TGCATTAAGCGAAGTTTGTAAATATTTTGGTGAAGAAATGAAGCTCATGCTGCACT 4423  
 5044 ATATGATGATATTTCTTAATGTTGAAGTTCACAAAGATTAATGATTTTGTGA 5103  
 4424 ATATGATGATATGTTTAACTGCTGAATATTAATTTCTTAACGATTTCAAACT 4483  
 5104 TCAATTTGAGATCACTTTGAAGTTAAAGTGTGGTGAAGAAATATCAATTTCTGTTAT 5163

4484 TGAATCGCTAAACATTTGCACTTAATATTTGCTGACATATCTGAATTTCTTAGAAT 4543  
 5164 TGAATTTGTAACCGGAATCTGTTATATTTTATCTCAAGAAAAATTTCTCAAGAAAT 5223  
 4544 TGAATTCATTAATTTGACAGGGGATATGATATATCAACAACAAATTTCTTGAATCTGT 4603  
 5224 ACTTAAGATTTGAACTAGATACATCATATGGAAGAAAAATATCCCTGATTCGCAATGA 5283  
 4604 AATTTAAAAATTAATTAATTAATTTTGAAGATATATTCATTAATTTTAAAG 4660  
 5284 CAATATGAAGAGTTGCAATTAATTTGCTGAAAAAGTTTATCCAGAGATGATTTTGAAA 5343  
 4661 GAAAAATTAATTTCCGAAA-----CTCACTAAAGAGTAATTTACTAATACAGAAAT 4714  
 5344 GGTTCGAAATGACATTTGCTTACCTGATGCTTAAAAAATTTTACCAAGAGTGTGG 5403  
 4715 CATTAATAAATGAATCATTTGCTTAATGAAGAAATTAAGCTATATCAATCAAGTGTGG 4774  
 5404 CTGCTTTTATGGGCTGCCAACAACAGCTCCAGATATATCGGTGATGATTCGTT 5463  
 4775 TTGCTTATTTAGGCACTAATTAACAAGGCGCCAGACATAGTTTGTCTGTAATCAAT 4834  
 5464 GGGTTTAAATCTGCAATTCAAATGCTCATGATTAAGAAATTTGATTTTGTCTTGA 5523  
 4835 GAGCTCAATTAATCAAAATCCGACATGATGATCTTGAAAAATTTGATATATGTTTACG 4894  
 5524 GTATATCAAAATATAGATGGATATATCAATTTGATGATCAAAAGAAACAGATTAATATCC 5583  
 4895 TTATGTAACAAACAAATATCTTTTATGATTAAGATTAAGCGAATCGATTTTCAATTA 4954  
 5584 ACCAAATATCTTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5643  
 4955 AAGGATCATTTATCATTTCAAACTTTTCAATGATCTTTTGCACCCCAAAAGATG 5014  
 5644 AATATCTATTTAGGCACTTTGATTTATGATGAATGAATTTGCTGCAATGGCGCCAA 5703  
 5015 AAGATGATTAACGTTATATACATTTTAAATGATTAATTAATTTGATTAATTTGCTACAT 5074  
 5704 AAAAAAAGCGTATAGCAACAAAGCTCAGCAGCTTGAATGTTGCTCTAATTTATAC 5763  
 5075 AAGGAGAGGATATCACTGATGATGTTCTGACGCTTGAAATCAAGCATCATTTACAG 5134  
 5764 AATGTTGAAGCTATGAAATTAATAAATCAATTTATGATTTGGGTTTGAAGTACGTA 5823  
 5135 TGTGAAGATGACCTTAATCAAGACACATATTTAGATCTAAATTTGATTAATGATGA 5194  
 5824 GATACATTTGATCAAGACAAACCAAGCTGATTAAGTTTGAAGAAATTA---CTATTG 5880  
 5195 AATTAATATTTTGAAGACAAAGCCGAGTTATAGGAATCTGTAATTAAGAGTACTTC 5254  
 5881 TCAACCAATGACCAATAGATATCTGATTAAGTTTCTAGCCCAATTTGATGATTA 5940  
 5255 ATATTAAGCGGATGCTGATATCAATTAATTAATTTATCAAGAGTTGTTGGAAG 5314  
 5941 AGATATTTCAATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 6000  
 5315 TATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5374  
 6001 TCTAATGCTGCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6060  
 5375 ATTAAGCCGAAATTAATTTGAAGAAATCTTGTCTACATATTTGAAGAAATGATTA 5434  
 6061 AGATATTAACACTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6101  
 5435 TAAAGATAGATCAATTTATGATGATGATGATGATGATGATGATGATGATGATGATG 5475

RESULT 6  
 AF078809  
 LOCUS  
 DEFINITION  
 ACCESSION

AF078809 1470 bp DNA 1linear PIN 03-AUG-1998  
 Candida albicans Tca4 retrotransposon reverse transcriptase (pol)  
 gene, partial cds.  
 AF078809



VERSION AF078809.1 GI:3377678  
 KEYWORDS  
 SOURCE Candida albicans  
 ORGANISM Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; mitosporic Saccharomycetales; Candida.

REFERENCE 1 (bases 1 to 1470)  
 Goodwin, T.J.D.  
 Teas, a Ty1-copia retrotransposon from Candida albicans  
 Unpublished  
 2 (bases 1 to 1470)  
 Goodwin, T.J.D.  
 Direct Submission  
 Submitted (16-JUL-1998) Department of Biochemistry, University of Otago, Cumberland Street, Dunedin, New Zealand

FEATURES  
 source  
 1..1470  
 /organism="Candida albicans"  
 /mol\_type="genomic DNA"  
 /strain="SC5314"  
 /db\_xref="taxon:5476"  
 <1..>1470  
 /gene="pol"  
 /gene="pol"  
 /codon\_start=2  
 /transl\_table=12  
 /product="reverse transcriptase"  
 /protein\_id="AAC28090.1"  
 /db\_xref="GI:3377679"  
 /translation="GRSVHPKRYLNAIVKIDYNCKEHPQSEEEIEKRNQVYTV  
 EKPKNVPLKTMVHTYKTMIDLNHNYKRCVYGVNMYENEDPDIHSSVPLT  
 SIKLSAIVANNIWMHOLDIASVLAASLEDGVIIVPRGGEVPGVSMILHSV  
 YGLQSAHNNVSHKRVNLEANGIKOTLHNGIIPKYNENDVIYVAYVDPMIATN  
 EKIIKEFAMLETFQIQYREBATEYIGIQRKRPDYITDQIPLEKVIATENIDS  
 YGKQIPILPKDINVVKQLRKSSQINDVKEKPKQLINSVKTIYDIEHEHPK  
 ESPQAEPLSAIGSLYQSAVGSILMASMTRPDLAFSPNOLGAKCAPDVDMKRLVY  
 CLRYKMDKLEKLEKRGRLNNKSDPIIBCFSPASPADLDRKSIIGTSIFVNGMLV  
 AMATKOKIITHSSAACEMLANIYTIKAFDLRTIEDL"

repeat\_region  
 <1..>1470  
 /note="Tca4 retrotransposon"  
 /rpt\_family="Ty-1-copia"  
 /rpt\_type="dispersed"

ORIGIN  
 Query Match 5.3%; Score 339.2; DB 8; Length 1470;  
 Best Local Similarity 54.4%; Pred. No. 3.4e-44;  
 Matches 790; Conservative 0; Mismatches 578; Indels 84; Gaps 2;

4423 TGGTAGAAGCGTTATTAACCCCAAGAGGTATTACATGCGGTCTGCAACAAATCGACTA 4482  
 1 TGGAGATCAGTGCATCCACCTAAAGATATTAAATGCCATTGTAAAGAAATAGATTA 60  
 4483 TAAATGCCGAGTAGATTAAGTCAATGATGCTGAAGTGAAGAAATTAGATCAAGA 4542  
 61 CAATCAAAAGATGCGCTCAAGTATGAGAGAGAAATCGAAATTTTAGGCTAACCA 120  
 4543 TCTTTACGAGAGTTCATTCACCGCGGTGAAGCCTATATCTATGAGGTGGGTACA 4602  
 121 AGTTTACACCGTTGAAAAACCAAGACGTCGTCCATTGAAAAACATGAGGTGACA 180  
 4603 TACTGAGAAATTTGATCTCTCAAGGTGTCTTCGGAATCAGCTGTGTTCACATG 4662  
 181 TACTTACAAACCAATGACCTCAAAATCATATTAACAAAGCGGTGGGTGATAGGG 240  
 4663 CAACAGACAAAGAAATTTGATGACCCCTTTAGTGTAGTTAGTCACTGTATAGA 4722  
 241 AAATCATATGTCGAAAAATCGTATTTGATCCCATGCCATCTCCCGGTAGAGA 300  
 4723 TTTTGACTATTAAGATTATTGCAATATAGGTGTGAATTAGGAATGAATTTACAA 4782  
 301 TCTCAAGATATACGACTATTATCTGCATAGCTGTGAAAAATACTTGTTATGACCA 360

4783 TTTAGCGTAGTGGCGGTATCTAAATGCTCTATTACTCATTCATCAATTTATGT 4842  
 DB ATTGACATCGCCTCAGCTTATTGGAAGCCAGTTTGAGAGATGGAAGAAATCTTGT 420  
 4843 CTTTCCTCTAATCAGTACCTTTGAAAGAAACCATTTGTTGTTATGAAACGTTCTGT 4902  
 DB GAGACCAACCGGTGTTTGAAGTTAAACCTGCTATAGTTGGGTTTACCAAGTCTGT 480  
 4903 CTATGGGTTAAACAGTCGGGTTTGAATGGTATCACACATCAACAAAGAGATTGAGA 4962  
 DB GTACGGTCTTAGGACAGAGTCCCATTAATTTGATCTACATTTTAAAGATGTTGAGCC 540  
 4963 CATTGTTTACTCAAGTTTACACAAATGATGTTTATTCACATTAATATGAAAGGG 5022  
 DB AAATGTTTAAACAAACACTACCAATGATGCTTTTGAAGAAATATGAAATGG 600  
 5023 ATCAGTAATATTTAGTTAGTTATATGTTGATGATATCTTATGGTGAAGTTACAAA 5082  
 DB AGATGTATTAATGAGTGTATATGAGAGATGTTTATGATAGCTAATATGAAAA 660  
 5083 AGTTATGATAATTTTGGATCAATGAGATCATTTGAAAGTTAAAGTTGGTGA 5142  
 DB GATTATTAAGAGTTTGTGTATGCTAGTCGAACATATTTCACTACATATTTGGCGA 720  
 5143 AATATCAATTAATCTTGGTATTTGAATTTGTAACCAATCTGTTATATTTATCTCA 5202  
 DB GGTACGAGATCTAGGAATATACATTGAAAGACACCTGATGATACAGTTAGACA 780  
 5203 AAAAAATTTCTCAAGAAATTAATTAAGATTTCAACTAGATACATATGAGAAAA 5262  
 DB AATACCATTTCTGAAAAATTTGGTCGTACATTCATATTCAGATGTTATGAAAAA 840  
 5263 CATACCCTGATTCGAATGACAAATATG----- 5321  
 DB CATTCCTATATCCCAAGAAATATTAATGTGTCAAGCAATTGAAAGTCAAGTCAAT 900  
 5292 AAAAGTTGCATATTTGTGTAAC----- 5317  
 DB AATGATTTCTGTAAGTGAGAAACCTTCAAAAGTTAATCAATTCGTTGCTCAAAACCA 960  
 5318 -----GTTAATCCAGAGATATTTTGAAGTTCCGAATGAGAC 5358  
 DB GTACATGATATGATACGAAGAAATGAGAAACATGATTTTAAAGATCACCAAGCTGA 1020  
 5359 ATTGCTTGAACCTGATGCTAAAAAACTATACCAAGTGTGGCCGCTTTATGGGC 5418  
 DB ACCCTTAAGCGGAAAGAAATTAATTAATACCAATCCGAGTGGACGTGTTATGAGCC 1080  
 5419 TGCACAAACACAGCTCCAGATATATCGGTGCTAGTGAATTCGTTGGTTCTAAATCTGC 5478  
 DB TTGATGAACACTAGGCCAGACTTAGCTTTAGTGAATCAGCTCGAGCAAGGTGTC 1140  
 5479 AAATCCAAATGCTCATGATTAAGAAATTTATTTATGCTCTAGGTATATCAAAAAAG 5538  
 DB TCAACGAGATGTCGACGCTGAGAGATGATGATGATGCTTACGATTAATTAAGAAAGA 1200  
 5539 CATGGATATCATTTGATGTAACAAAGAAACAGATTTGAATATACCAACCAATCTTGT 5598  
 DB TATGGACTTCAAACTTGAATACAGAGAGAAAGTCAACAAAGTCAAAAGATTTTAT 1260  
 5599 TATGAAATGTTCAAGTATGCTCATTTGACACAGATTTGATGAAATCTTATTAAGTG 5658  
 DB TATGAAATGTTCTGAGTGCCTCATTTGCCCAATTTGACAGAAAGTCAATACCGG 1320  
 5659 AACTTGAATTAATGTAAGAAATTTGGTGAATGGGGGACCAAAAAACAAACGTCAT 5718  
 DB GACAAACATTTTGTAAATGGAATCTGTTGATGAGGCAACAAAGAAACAAAAATAT 1380  
 5719 AGCAAAAGCTCAGAGCTTGTGAATGTTGCTCTAAATTAATACAAATGTTGAAGCTAT 5778  
 DB CACACATATTCAGCGCTTGGGAATGCTTCATTAATTAATTAATTAATTAAGGCAAT 1440  
 5779 CGAATTAATAAA 5790



Db 1441 TGATTGAGAA 1452

RESULT 7  
LOCUS AX073190 1888 bp DNA linear PAT 25-JAN-2001  
DEFINITION Sequence 301 from Patent WO0102550.  
ACCESSION AX073190  
VERSION AX073190.1 GI:12583390  
KEYWORDS  
SOURCE  
ORGANISM Candida albicans  
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
Saccharomycetales; mitosporic Saccharomycetales; Candida.

REFERENCE  
AUTHORS Contreras, R.H., de Backer, M.D., Luyten, W.H., Malcorps, I.K.,  
Neijssen, B.J., and Reekmans, R.J.  
TITLE Cell death related drug targets in yeast and fungi  
JOURNAL Patent: WO 0102550-A 301.11-JAN-2001;  
JANSSEN PHARMACEUTICA N.V. (BE)  
FEATURES  
source 1..1888  
/organism="Candida albicans"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:5476"

ORIGIN  
Query Match 2.6%; Score 166; DB 6; Length 1888;  
Best Local Similarity 97.1%; Pred. No. 1e-16;  
Matches 169; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 115 TGTATATTAATTAAGGATGATGAATACCAATCCCAATATCAAGAGATAGAAAGGA 174  
Db 1 TGTATATTAATTAAGGATGATGAATACCAATCCCAATATCAAGAGATAGAAAGGA 60

QY 175 GGAAGTTTCAATATATCTTGGAATATTAATCTTGCTATTCATCTACATCACTAG 234  
Db 61 GGAAGTTTCAATATATCTTGGAATATTAATCTTGCTATTCATCTACATCACTAG 120

QY 235 ACGTGTACACGCTCATCTCAGGTAAAGAAAGTTTATTCATCAGATTAGAA 288  
Db 121 ACGTGTACACGCTCATCTCAGGTAAAGAAAGTTTATTCATCAGATTATTA 174

RESULT 8  
PFMAL13\_07  
WFCOMMENT  
Sequence split into 28 fragments LOCUS PFMAL13 Accession AL844509  
Fragment Name Begin End

PFMAL13\_00 1 110000  
PFMAL13\_01 100001 210000  
PFMAL13\_02 200001 310000  
PFMAL13\_03 300001 410000  
PFMAL13\_04 400001 510000  
PFMAL13\_05 500001 610000  
PFMAL13\_06 600001 710000  
PFMAL13\_07 700001 810000  
PFMAL13\_08 800001 910000  
PFMAL13\_09 900001 1010000  
PFMAL13\_10 1000001 1110000  
PFMAL13\_11 1100001 1210000  
PFMAL13\_12 1200001 1310000  
PFMAL13\_13 1300001 1410000  
PFMAL13\_14 1400001 1510000  
PFMAL13\_15 1500001 1610000  
PFMAL13\_16 1600001 1710000  
PFMAL13\_17 1700001 1810000  
PFMAL13\_18 1800001 1910000  
PFMAL13\_19 1900001 2010000  
PFMAL13\_20 2000001 2110000  
PFMAL13\_21 2100001 2210000  
PFMAL13\_22 2200001 2310000

PFMAL13\_23 2300001 2410000  
PFMAL13\_24 2400001 2510000  
PFMAL13\_25 2500001 2610000  
PFMAL13\_26 2600001 2710000  
PFMAL13\_27 2700001 2732359  
Continuation (8 of 28) of PFMAL13 from base 700001 (AL844509 plasmidium falciparum 3D7 c

Query Match 1.7%; Score 107; DB 2; Length 110000;  
Best Local Similarity 47.0%; Pred. No. 7.7e-08;  
Matches 332; Conservative 0; Mismatches 375; Indels 0; Gaps 0;

QY 3601 AGAAGATTGGCATCTCTTGACATCCCAAACTAGTGTGAAGATGACACAGCAAAATCC 3660  
Db 42996 AATGATTAATTAATTAATGAATGAATGATATATATATATGATGATTAATTAATGA 43055

QY 3661 CTCGTACTGGCTAATTAATGAATGCGGTAACTGAAAAATAGAGAGGAGAAATTTTC 3720  
Db 43056 TGATGATGATTAATTAATGATGATGATGATATATATATATGATGATGATTAATTA 43115

QY 3721 ATTTCGGGGGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3780  
Db 43116 TGATGATGATGATTAATTAATGAATGATGATGATGATGATGATGATGATGATGAT 43175

QY 3781 TGAACAGAGAGATGCTGTAAAGTCCCAATTCAGACGAAAGTTCTCAAGAGGAGAAAT 3840  
Db 43176 TAAATGATGATGATTAATTAATGATGATGATGATGATGATGATGATGATGATGAT 43235

QY 3841 ACTTAATGAACAACTGATTAATGTTGATGATGATGATGATGATGATGATGATGATGAT 3900  
Db 43236 TAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 43295

QY 3901 CTCCTCTTAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3960  
Db 43296 TAAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 43355

QY 3961 CAATTCAGAGATCCGATTCATCGAATGAAATGAAATGATGATGATGATGATGATGATG 4020  
Db 43356 TGATTAATTAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 43415

QY 4021 AATTACTGAAATTAATCAAAAACACTGAAAGTATCTTCTGATTAAGATGATGATG 4080  
Db 43416 TGATGATTAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 43475

QY 4081 ATTGAAGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4140  
Db 43476 TGATGATTAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 43535

QY 4141 AGCTAATCCACTACGATGAAATTAATGATGATGATGATGATGATGATGATGATGATG 4200  
Db 43536 TGATGATGATGATTAATTAATGATGATGATGATGATGATGATGATGATGATGATG 43595

QY 4201 TCATACAGTGCACGAGCAAGAAAGAAATACATTGGAAGAAAGTTAAGGAAATAT 4260  
Db 43596 TAATTAATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 43655

QY 4261 TGTGTGACTAAACATTCATGATTAACAATGATCAATGATCAAGAGAGCA 4307  
Db 43656 AGAAGATCATTTAATTAATGATGAAACCCCAATGATTAAGAGACTA 43702

RESULT 9  
AC140550 128101 bp DNA linear HTG 12-DEC-2003  
LOCUS Medicago truncatula clone mh2-54a24, WORKING DRAFT SEQUENCE, 2  
DEFINITION  
ACCESSION AC140550  
VERSION AC140550.19 GI:39752730  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE Medicago truncatula (barrel medic)  
ORGANISM Medicago truncatula  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Medicago.  
1 (bases 1 to 128101)  
Shauli, S., Lin, S., Dixon, R., May, G., Summer, L., Gonzales, B.,  
Cook, D., Kim, D. and Roe, B.A.  
Medicago truncatula BAC Clone mch2-54a24  
Unpublished  
2 (bases 1 to 128101)  
Shauli, S., Lin, S., Dixon, R., May, G., Summer, L., Gonzales, B.,  
Cook, D., Kim, D. and Roe, B.A.  
Direct Submission  
Submitted (25-FEB-2003) Department Of Chemistry And Biochemistry,  
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
OK 73019, USA  
3 (bases 1 to 128101)  
Shauli, S., Lin, S., Dixon, R., May, G., Summer, L., Gonzales, B.,  
Cook, D., Kim, D. and Roe, B.A.  
Direct Submission  
Submitted (12-DEC-2003) Department Of Chemistry And Biochemistry,  
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
OK 73019, USA  
On Dec 12, 2003 this sequence version replaced gi:38678580.  
----- Genome Center  
Center: Department Of Chemistry And Biochemistry  
The University Of Oklahoma  
Center code: UOKNOR

NOTE: This is a 'working draft' sequence. It currently  
consists of 2 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
This record will be updated with the finished sequence  
as soon as it is available and the accession number will  
be preserved.

1 14715: contig of 14715 bp in length  
\* 14716 14815: gap of unknown length  
\* 14816 128101: contig of 113286 bp in length.  
Location/Qualifiers  
1. 128101  
/organism="Medicago truncatula"  
/mol type="genomic DNA"  
/db xref="taxon:3880"  
/clone="mch2-54a24"  
/clone\_lib="Medicago truncatula BAC library H2"

ORIGIN  
Query Match 1.6%; Score 105.6; DB 2; Length 128101;  
Best Local Similarity 47.3%; Pred. No. 1.2e-07;  
Matches 352; Conservative 0; Mismatches 389; Indels 3; Gaps 1;

4470 ACAAAATGACCTATATGATCCCGGATGATTAAGTCAATGATGCTGAGAAAT 4529  
69396 ACAAAAGACATGAGACTAATGTTGAAAAGTCTATGTTGAAGAACTACATCAA 69455  
4530 TTAGATCAAAAGATGTTACGAGAAAGTTCCCATTTCCACCGGTGTAAGCTATATCTA 4589  
69456 TTGAGAGAAATCAAACTGAGAGCTAGTTAAGTCCGACAGAAAAGAAAGATGATG 69515  
4590 TGGGTTGGTACATCTGAGAAATGATTTCTTCAAAAGTCTGTTCCGAAAACAGCTT 4649  
69516 TCAAAATGGGTTTTCAGAGTGAATTTGAATCTGATGTAAGTTCTTAAGCAAAACAA 69575  
4650 GGTGTTGTCATGAGCAAGACAAAGAAATGATGATGATGATGATGATGATGATGAT 4709  
69576 GATTGGTGTAGAGGTTTCTTCGAAAGCAA---GACTTGAATTAAGTAAGTGTGTTG 69632  
4710 CACCTGTTATGATCTTGTGACTATTAATTAATGACATATATAGTTTGAATTAGAA 4769  
69633 CTCGAGTTGCTAGAAATGAATGAACTGTGAGGTTAGTAAGCTATGCTGTGAAGATGAAT 69692  
4770 TGACAAATCAACATTAGACGCGTGAAGGCGATCTTAATGCTTATTAATCAATCAA 4829  
69693 GGTCACTCTACCATTTAATGATGTAAGTACGATTTCTTAATGCTCCACTAAGAGGTAG 69752

QY 4830 ATCCAAATTATGCTTTTCCCTCAATACAGTACCTTTGAAGAAAACCAATGTTGTTAT 4889  
DB 69753 TGTGTTCTCAACACCAACCAATTTGAAATGCTGGAGAAAATATGTTGATTAAC 69812  
QY 4890 TGAACGTTTGTCTATGAGTTAAACAGTGGGTTTGATGATCACTATCAAA 4949  
DB 69813 TGCAAAAACACTATGATTAATAATGAGTCTTGAAGCATGGAACGAAATTAATC 69872  
QY 4950 GAGTATGGAAGACATGTTTCTTACTCAAGTTTACAAATGATGTTTATTTTCACATG 5009  
DB 69873 AAGTTTATCAATTAATGAGTTTAAATTTGATGATGATGATGATGATGATGATGAT 69932  
QY 5010 AATATGAAGAGGATCACTATATATTTAGTTTATATGTTGATGATATTTTATGTTG 5069  
DB 69933 AGCTTATGATGAGAGAACTGTTATTAATTTGCTGATGATGATGATGATGATGATGAT 69992  
QY 5070 GAGTTCACAAAAGTTATGATATTTTGGATCATGATGATGATGATGATGATGATGATGAT 5129  
DB 69993 GTAGTCAACATCAGAGTTTGAAGAGTTTAAAGAAACATGAATGAGTTGATTAATGA 70052  
QY 5130 AAGTGTGTTGTAATATCAAAATTAATCTGTTGATGATGATGATGATGATGATGATGAT 5189  
DB 70053 CTGATCTTGGGAGCTTCATTTCTTGAAGAAATGATGATGATGATGATGATGATGATGAT 70112  
QY 5190 ATATTTATCTCAAGAAAATTTTC 5213  
DB 70113 TAGTATGATATCAAGAAAATTTTC 70136

RESULT 10  
AL935272 146275 bp DNA linear HTG 30-OCT-2003  
AL935272 Damio rerio clone CH211-266K2, WORKING DRAFT SEQUENCE, 2 unordered  
DEFINITION  
pieces.  
AL935272.16 GI:37936458  
VERSION  
HTG: HTGS PHASE1; HTGS ACTIVERIN; HTGS\_DRAFT; HTGS\_FULLTOP.  
KEYWORDS  
Damio rerio (zebrafish)  
SOURCE  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Osteichthyes;  
Cypriniformes; Cyprinidae; Danio.  
1 (bases 1 to 146275)  
Andrew R.  
Direct Submission  
Submitted (29-OCT-2003) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SN, UK. E-mail enquiries:  
zfish-help@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk  
On Oct 23, 2003 this sequence version replaced gi:35763272.  
----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: zfish-help@sanger.ac.uk  
----- Project Information  
Center project name: zc266K2  
----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Chemistry: Dye-terminator; 100% of reads  
Consensus quality: 145872 bases at least Q40  
Consensus quality: 145995 bases at least Q20  
Consensus quality: 146056 bases at least Q20  
Insert size: 146175; sum-of-contigs  
Insert size: 155501; 1.5% error; agarose-fp  
Quality coverage: 9.39x in Q20 bases; sum-of-contigs Quality  
coverage: 9.13x in Q20 bases; agarose-fp

NOTE: This is a 'working draft' sequence. It currently  
consists of 2 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.

*Lotus corniculatus* var. *japonicus* (*Lotus japonicus*)

3671 GCTAATAGCATCGGTA CTGAAAAATAGATGAGGAGAAATATTTCATTTCGGGG 3730

FEATURES	REFERENCE	LOCUS	LOCUS DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
		AP006375/c						
		AP006375	72880 bp DNA linear					PLN 22-JUL-2003
			Lotus corniculatus var. japonicus genomic DNA, chromosome 4, clone:LFJ32M04, TM0219, complete sequence.					
		AP006375						
		AP006375.1	GI:31581006					
		HTG.						
			Lotus corniculatus var. japonicus (Lotus japonicus)					
			Lotus corniculatus var. japonicus					
			Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Lotaeae; Lotus.					
		1						
			Aamizu, E., Kato, T., Sato, S., Nakamura, Y., Kaneko, T. and Tabata, S.					
			Structural Analysis of a Lotus japonicus Genome. IV. Sequence Features and Mapping of seventy-three TAC clones which cover the 7.5 Mb Regions of the Genome					
			DNA Res. (2003) In press					
		2	(bases 1 to 72880)					
			Sato, S.					
			Direct Submission					
			Submitted (07-MAY-2003) Shusei Sato, Kazusa DNA Research Institute, Department of Plant Gene Research, 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 259-0818, Japan (E-mail: ssato@kazusa.or.jp), URL: http://www.kazusa.or.jp/, Tel: 81-438-52-3935(ex. 2337), Fax: 81-438-52-3934)					
			Location/Qualifiers					

FEATURES	Location/Qualifiers
source	1. .72880

Location/Qualifiers  
1. . 72880

```

/organism="Lotus corniculatus var. japonicus"
/mol_type="genomic DNA"
/variety="japonicus"
/db_xref="taxon:34305"
/chromosome="4"
/clone="UJT32M04"
/clone_1b="UJT_library"
/note="TAC clone:TW0219-synonym: Lotus japonicus"

```

ORIGIN

/note="TAC clone: TM0219~synonym: Lotus japonicus"

Query Match 1.6%: Score 105.4: DB 8: Length 72880:

Match 1.68; Score 105.4; DB 8; Length 72880;

Best Local Similarity 47.0%; Pred. No. 1.6e-07;  
Matches 386; Conservative 0.7; Mismatches 441;  
Indels 6; Gaps 2.

Local Similarity 47.08; Pred. No. 1.6e-07;  
196; Conservative 0; Mismatcher 441; Indels 6; Gaps 3;

Machine	Conservative	MisMachine	File	Index	Days	2
Machine	Conservative	MisMachine	File	Index	Days	2

Year	Conservative	Liberal	Other
1980	45%	45%	10%
1984	48%	42%	10%
1988	52%	38%	10%
1992	55%	35%	10%
1996	58%	32%	10%
2000	60%	30%	10%
2004	62%	28%	10%
2008	65%	25%	10%
2012	68%	22%	10%
2016	70%	20%	10%
2020	72%	18%	10%

44/1 CAAATCGACTATTAATGATCCGGATGGTAAAGTCAATGAAAGCTGAACTAGAGAAAT 4530

44/1 CAAAATCGACTAAATGATCCGGGATGGATAAAGTCAATGGAATGCTGAACTAGAGAAAT 4530

D5 5653 CAAGAGCTCTGACTATGAGTTCTGGATTCAATCCATGCAGAGGGAACGGGACAGTT 5647

56530 CAAAGAGCTCTGACTATAGCTTGGATTCAATCCATGCAAGAGGAATTGGACAGTT 5647

4531 TAGATCAAAGATGTTTACGAGAAGTTCGAATTCACCGGTGTGAAGCCTATATCTAT 4590

4531 TAGATCAAAAGATGTTTACGAGGAAGTTCCTCAATTCCACCGGTGTGAAGCCTATATCTAT 4590

Db 56470 CAGAGGATGACTGTGGGATTTGTTCCAGACCTGATGATGCAATGTTGTGGGAA 5641

56470 CAAGAGCATGAGTGTGGGATTTGGTCCAGACCTGATGATGCACAATGTTGTGGGAAC 5641

4591 GGGTTGGTACATCTGAGAAATTGATTTCTCTCAAAGCTGTTGTCGGAATCACGTTG 4650

4591 GGGTTGGTACATACCTGAGAAATTGATTCTCTCAAGGTGTTGTTGGAAATCACGTTG 4650

Db 56410 TAAGTGATATTCAGCAACAAGTCTGATGAAGTGTATGTGACACAAGAACAAATCTCG 5635

56410 TAAGTGATATTCAGGACACAGTCTGATGAAAGTGGTAATGTGACAAAGGACAAATCTCG 5635

4651 TGTTCATGGCAACAGACAAAAGAAAAATTGGATTATGACCTTTAGTGTAGTTC 4710

4651 TGTGTCCATGGCAACGACAAAGGAAAAATTGGATTATGACCCCTTTAGTGTAGTTC 4710

dbs  
56250 TCTTACT TCCTCAGGCGATTACTCTCAATAACAAATGAACAATTCATTTCTC 5629

TTCCTCAGGCGATTACATCCAAATAACAAGCAATGCATGGAAACTTTTGC

66350

66359

[illegible][illegible]

411ACCGTATAGATCTTGACGTAAGGATTA1GACCAATAAGGTGGAATAGGAAT 41/10

4/11 ACCGCTATAGACTCTGACACATAMGATTAATGACCAATAAAGCTGTGAAATAGGAAAT 4/10

D6 56293 TCCTGTGCTCGCTCTTGATCTATCAGGTTGTTGTTGGGAGTACGATGTTGCTAAATT 5623

56293 TCCTGTGCTCGCTTGAATCTATCAGGTTGTTGTTGGGAGTACCATGTTGCTAAATT 5623

Qy 4771 GACAAATCAACATTTCAGCGTCGAGTCGGCGTATCTAAATGCCCTCATTTACTCATTTCAA 4830

4771 GACAAATTCACATTAGACGCTGAGTCGCGTATCTAAATGCCCTCTAATTACTTCATTCAAA 4830

Db 56233 CAGACTATATCAGATGGATGTCAAGAGTCATTCTTGAATGGTTACTTGAATGAAGAAGT 5617

56233 CAGACTATATCAGATGGATGTCAGAGTGCATTCTTGATGGTTACTTGATGAAGAAGT 5617

4831 TCCAATTATGCTCTTCTCTCTAATCAGTACCTTTGAGAAAAACCATTTGTTATT 4890

4831 TCCTATTATGCTTTCCTCCTAATCAGTACCTTGAAGAAAAACCATTTGGTTATT 4890

Db 56173 CTATGTGAACAGCCTAAGGTTTGTAGATCCAGTTTCCAGATCATGTATACAGATT 5611

56173 CTATGTGAACAGCCTAAAGGTTTGTAGATCCAAAGTTTCAGATCATGTATACAGATT 5611

4891 GAAACGTTCTGTCTATGGGTTAAACACGTGGGTTTGGATGGTATCACACTATCAAAG 4950

4891 GAAACGTTCTGTCTATGGGTTAAACAGTCGGGTTTGGATGGTATCACACTATCAAAG 4950





\* 11700 29732: contig of 18033 bp in length  
 \* 29733 29832: gap of unknown length  
 \* 29833 129511: contig of 39679 bp in length.

FEATURES  
 Location/Qualifiers  
 1..129511

/organism="Medicago truncatula"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:3880"  
 /clone="mch2-1036"  
 /clone\_lib="Medicago truncatula BAC library H2"

ORIGIN

Query Match 1.6%; Score 103.4; DB 2; Length 129511;

Best Local Similarity 46.9%; Pred. No. 2.7e-07;  
 Matches 391; Conservative 0; Mismatches 436; Indels 6; Gaps 2;

4471 CAAATGCACTAATATATCCGGATGATTAAGTCAATGAAGCTAGAGAAATT 4530  
 Db 50154 CAAAGAAAGCCCTAAGTGTGATTTGGATTGAAAGCAATGCAATGAAGCTAGATTT 50213  
 Qy 4531 TAGATCAAAAGATGTTTACGAGAAGTTCATCCACCGGTGAGGCTATATCTAT 4590  
 Db 50214 CAAAGAAAGTGAAGTGTGATTTAGTTCCAGAGCCGAGGCTATATGATGATAGGTAC 50273  
 Qy 4591 GGGTTGGTACATCTGAGAAATTTGATCTCTCAAGGTGTGTTCGGAATCAGCTTG 4650  
 Db 50274 CAAAGTGAATCTTAAAGAAATCTGATGAA--AATGTAATCTGTAACGAGAAACAGGC 50330  
 Qy 4651 TGTTCATGCAACAGACAGAAAGAAATTTGATTAAGCCCTTTAGTTAGTTTC 4710  
 Db 50331 AAGATTATAGCTTAAGCTACACATATGAAGGTCTGACCTTTATCAGACTTTGC 50390  
 Qy 4711 ACTGTTATAGATCTTGTGACTATPAGATTATGACATATATAGGTTGTGAATTAGCAAT 4770  
 Db 50391 TCTGTTGCTGATTTGAATATATAGACTATGCTGATGATATCATGATATACCAAGTT 50450  
 Qy 4771 GACATTTCAACATTAGACGTCGAGTGGCGCTATCTAAATGCTCTATTTCTCATCAAA 4830  
 Db 50451 TAAGTTCTTCAATAGATGTAAGAGAGCTTTTGAATGGTACTGAAAGAGAGAGT 50510  
 Qy 4831 TCCATTATATGCTTTCTCTTAATACGATACCTTTGAAGAAACCATTTGGTTATT 4890  
 Db 50511 CTATGTTAGCAACCTAAGAGGTTATATGATCTAGCTTCCCAATCATGTATAGATT 50570  
 Qy 4891 GAAAGCTTCTGCTATGTTGTTAAAGAGTGGGTTTGAATGATACACTATCAAAAG 4950  
 Db 50571 GAAAGAGCTTGTATGCTTAAAGCAGGCACTTAGGCTTGTATGATGACTCTCA 50630  
 Qy 4951 AGTATGGAAGACATTTGTTTACTCAAGTTTACACATGATGTTTATTTCACTTGA 5010  
 Db 50631 GTTTATTTGTTGTCAGGCTACATGAAGAGATACATACAGACCTTGT---GTAAA 50687  
 Qy 5011 ATATGAAGAGGATAGTAATATATTTAGTTTATGTTATGATATTTCTATAGTTGG 5070  
 Db 50688 AGATGACAAAGAGACATGATGATGCTCATATATATGTTGATACATGTTTTTGTGG 50747  
 Qy 5071 AAGTTCACAAAAGTTATGTAATTTTGTGATCAATGAGATCATTTTGAAGTTAA 5130  
 Db 50748 AATGTCACAAAGATGTCACAACTTATTCAGAAATGCAATGATGATTTGTGATGAT 50807  
 Qy 5131 AGTGTGGTGAATATATCAATATATCTTTGATTTGATTTGTTAAACCAATGTGTTA 5190  
 Db 50808 CTATGCTGGTGTGAGCTTATTTTCTTGGCTTCATGTTTAAACAAATGAGAGTACTAT 50867  
 Qy 5191 TATTTATCTCAAGAAATTTCTCAAGAAATTTCTTAAGATTTCAAAGTATGCTC 5250  
 Db 50868 CTTCATTTCTCAAGTATATATGTCACAGACATGATGAGAAATTTGGCTTGAATATG 50927  
 Qy 5251 ATATGGAAGAAACATACCTGATTCGAAATGACAAATATGAAAGTTGCA 5303  
 Db 50928 TGTGATATAGAAAATCCCAAGTGCATTTGTAAGCTGAAAAAGATGAAA 50980

RESULT 15  
 AC135396 134544 bp DNA linear HTG 13-DEC-2003  
 LOCUS AC135396/c  
 DEFINITION Medicago truncatula clone mth2-33018, WORKING DRAFT SEQUENCE, 2  
 uncloned pieces.

AC135396  
 AC135396.25 GI:39841085  
 VERSION HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
 KEYWORDS Medicago truncatula (barrel medic)  
 SOURCE Medicago truncatula  
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eustoside I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;  
 Medicago.

REFERENCE  
 1 (bases 1 to 134544)  
 Shaull, S., Lin, S., Dixon, R., May, G., Sumner, L., Gonzales, B.,  
 Cook, D., Kim, D. and Roe, B.A.  
 Medicago truncatula BAC Clone mth2-33018

REFERENCE  
 2 (bases 1 to 134544)  
 Shaull, S., Lin, S., Dixon, R., May, G., Sumner, L., Gonzales, B.,  
 Cook, D., Kim, D. and Roe, B.A.  
 Direct Submission

TITLE  
 JOURNAL  
 Submitted (14-OCT-2002) Department Of Chemistry And Biochemistry,  
 The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
 OK 73019, USA

REFERENCE  
 3 (bases 1 to 134544)  
 Shaull, S., Lin, S., Dixon, R., May, G., Sumner, L., Gonzales, B.,  
 Cook, D., Kim, D. and Roe, B.A.  
 Direct Submission

TITLE  
 JOURNAL  
 Submitted (13-DEC-2003) Department Of Chemistry And Biochemistry,  
 The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
 OK 73019, USA  
 On Dec 13, 2003 this sequence version replaced gi:39752708.

COMMENT  
 ----- Genome Center  
 The University Of Oklahoma  
 Center code:UOKNOR

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 2 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

\* 1 2202: contig of 2202 bp in length  
 \* 2203 2302: gap of unknown length  
 \* 2303 134544: contig of 13242 bp in length.

FEATURES  
 source  
 1..134544  
 /organism="Medicago truncatula"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:3880"  
 /clone="mth2-33018"  
 /clone\_lib="Medicago truncatula BAC library H2"

ORIGIN

Query Match 1.6%; Score 102.2; DB 2; Length 134544;

Best Local Similarity 49.5%; Pred. No. 4.2e-07;  
 Matches 263; Conservative 0; Mismatches 266; Indels 0; Gaps 0;

Qy 4724 CTGTGCACTAATAGATTTATGCAATATATAGTGTGATTTAGGAATGACATTTCAACAT 4783  
 Db 30371 CTAGAAAGCTGAGGTATGATGATCTATGCTTTGAAGATGATGACCACTACACAT 30312  
 Qy 4784 TTAGAGCTGATCGGCGGATGTAATGCTCTATATCTCACTTCAATCCAAATTATATG 4843  
 Db 30311 TTACATGTAAGTGCACATTTTAAATGATGCTGTAAGAGAGTGTGTTGTGTGACAA 30252  
 Qy 4844 TTTCCTCTTAATGATGATCTTTGAAGAAACCATTTGTGTTATGAAAGCTTTCTGTC 4903

Db 30251 CCACCAAGTTTGAATTCAGGAGGAAAAATGTGTATAACTGCATPAAGCACTA 30192  
 QY 4904 TATGGTTAAACAGTCGGGTTTGGANGGTATCACTATCAAAAAGATATGGAGAC 4963  
 Db 30191 TATGATTAAACAGGTCCTGAGACATGGAACAAAGAAATTGATCAAGTTTAAATCA 30132  
 QY 4964 ATTGTTTACTCAAGTTTACACATGATGTGTTTATTCACATTGAATATGAAGAGGA 5023  
 Db 30131 ATAGGCTTTAAAGATGTCAGTTGATGTATATGTTTCAGAAAGCCTAGATGA 30072  
 QY 5024 TCAGTAATATATTTAGTTTATATGTTGATGATTTCTTATGTTGAAAGTTCACAAAA 5083  
 Db 30071 GGAAGTGTATATATTTGTCGTATGTTGATGCTTACTCATCAGTAGCTCAACATCA 30012  
 QY 5084 GTTATTGATATTTTGTGATCAATTGAGATCAATTTGAAGTTAAAGTGTGGTGA 5143  
 Db 30011 GAGATTAAAGAGTTAAAGAGAACTGAAGCTGAGATTGAGATGATGATCTTGGTGAG 29952  
 QY 5144 ATATCAATTAATCTGTGATTTGTAATTTGTAACCGAATCTGTTATATTTATCTCA 5203  
 Db 29951 CTTTCATTCTTCTTGAAGATGAAATTTGAAATTTGAAGTTGGATAGTATATCA 29892  
 QY 5204 GAAAAATTTCTCAGAAATTAATTAAGATTTCAAACTAGATGACTCATAT 5254  
 Db 29891 AAAAATACATGTGTGAATGCTTGAATTAATTGAATGAAGAGCTGCAT 29841

Search completed: April 11, 2004, 18:29:14  
 Job time : 15982 secs

*This Page Blank (uspto)*



RESULT 1	ID	AAAS7920	standard; DNA: 6426 BP.
XX	AC	AAAS7920;	
XX	XX	15-SEP-2003 (revised)	
DT	DT	11-OCT-2000 (first entry)	
XX	XX		
DE	DE	Candida albicans pcal retrotransposon.	
XX	KM	Retrotransposon; pcal; Tca2; Ty1; copia; long terminal repeat; LTR;	
XX	KM	gag gene; group antigen; polyprotein; pol; aspartate protease; integrase;	
XX	KM	reverse transcriptase; RNaseH; pseudoknot; readthrough translation;	
XX	KM	stop codon suppression; gene delivery; gene therapy vector;	
XX	KM	genetic vaccine composition; immunogenic; transgenic animal; ds.	
OS	OS	Candida albicans; strain h031042.	
XX	XX		
FT	FT	Key	Location/Qualifiers
FT	FT	LTR	1..280
FT	FT	/*tag= a	
FT	FT	1..6	
FT	FT	/*tag= b	
FT	FT	/*rpt_type= INVERTED	
FT	FT	/*note= "imperfect 6 bp repeat"	
FT	FT	120..125	
FT	FT	/*tag= c	
FT	FT	/*standard_name= "TATA box"	
FT	FT	185..190	
FT	FT	/*tag= d	
FT	FT	/*standard_name= "TATA box"	
FT	FT	201..206	
FT	FT	/*tag= e	
FT	FT	275..280	
FT	FT	/*tag= f	
FT	FT	/*rpt_type= INVERTED	
FT	FT	/*note= "imperfect 6 bp repeat"	
FT	FT	281..291	
FT	FT	/*tag= g	
FT	FT	/*note= "this site in the corresponding RNA is a primer binding site for reverse transcriptase-mediated minus-strand DNA synthesis"	
FT	FT	496..6103	
FT	FT	/*tag= 1	
FT	FT	/*product= "gag (group antigen)-pol (polyprotein)"	
FT	FT		

FT	readthrough protein"
FT	/note= "The gag-pol readthrough protein is the result of
FT	stop codon suppression mediated by pseudoknot formation
FT	in the mRNA"
FT	/codon=(seq:"ctg", aa:Ser)
FT	/note= "12 serine residues in the gag-pol readthrough
FT	protein are encoded by this non-standard Ser codon"
FT	/transl_except=(pos:1367..1372, aa:Glu)
FT	498..1372
FT	/tag= h
FT	/product= "Gag (group antigen) protein"
FT	/codon=(seq:"ctg", aa:Ser)
FT	/note= "5 serine residues in the gag protein are encoded
FT	by this non-standard Ser codon"
FT	1373..6100
FT	/tag= x
FT	/product= "Pol (polypotein), comprising aspartate
FT	protease, integrase, reverse transcriptase (RT) and
FT	RNaseH"
FT	/codon=(seq:"ctg", aa:Ser)
FT	/note= "7 serine residues in the pol protein are encoded
FT	by this non-standard Ser codon"
FT	1381..1455
FT	/tag= j
FT	/note= "The RNA corresponding to this region forms a
FT	pseudoknot, allowing gag ORF stop codon suppression and
FT	translation of the gag-pol readthrough protein"
FT	3455..3465
FT	/tag= l
FT	/note= "Polypurine tract 2 (PPT2)"
FT	complement(3455..3465)
FT	/tag= m
FT	/note= "This site in the corresponding RNA is a primer
FT	binding site for reverse transcriptase-mediated plus-
FT	strand DNA synthesis"
FT	6136..6146
FT	/tag= n
FT	/note= "Polypurine tract 1 (PPT1)"
FT	complement(6136..6146)
FT	/tag= o
FT	/note= "This site in the corresponding RNA is a primer
FT	binding site for reverse transcriptase-mediated plus-
FT	strand DNA synthesis"
FT	6147..6426
FT	/tag= p
FT	6147..6152
FT	/tag= q
FT	/rpt_type= INVERTED
FT	/note= "Imperfect 6 bp repeat"
FT	6266..6271
FT	/tag= r
FT	/standard_name= "TATA box"
FT	6331..6336
FT	/tag= s
FT	/standard_name= "TATA box"
FT	6347..6351
FT	/tag= t
FT	6421..6426
FT	/tag= u
FT	/rpt_type= INVERTED
FT	/note= "Imperfect 6 bp repeat"
XX	
PN	WO200026397-A1.
PD	11-MAY-2000.
PF	01-NOV-1999;
PR	30-OCT-1998;
PR	30-OCT-1998;
PA	(JANNC ) JANSSEN PHARM NV.

XX	Luyten WHML, De Backer MD, Nelissen BJM, Foulter RTM;
DR	WPI: 2000-265640/31.
DR	P-PSDB; AAB03126, AAB03127, AAB03128.
XX	Novel retrotransposon expression vectors useful for expressing an
PT	antigen, epitope or therapeutic agent, or detecting genes or the presence
PT	of Candida in a sample.
XX	Claim 12, Fig 2B; 204pp; English.
CC	The invention relates to novel retrotransposons from the yeast <i>Candida</i>
CC	<i>albicans</i> which have a copy number of 40-150, preferably 50-100 copies per
CC	genome. In particular, the invention relates to the novel <i>C. albicans</i>
CC	<i>Ty1/copia</i> retrotransposon pCal (AA57920), and to the integrated form of
CC	this retrotransposon, designated TCa2, and to the novel <i>C. albicans</i>
CC	hog1042 and has a copy number of 50-100 copies per cell. It comprises
CC	identical 280 bp long terminal repeats (LTRs) and two open reading frames
CC	(ORFs). The first ORF encodes a gag (group antigen) protein, and the
CC	second ORF encodes a polypeptide (pol) consisting of an apertase
CC	protease, integrase, reverse transcriptase (RT) and RNaseH. The gag and
CC	pol ORFs of pCal are in the same reading frame, separated only by a
CC	termination codon (TGA). Translation of the pol ORF occurs through the
CC	occasional readthrough suppression of the stop codon, which is mediated
CC	by the formation of a pseudoknot within the gag-pol mRNA. The
CC	retrotransposons of the invention can be used as vectors for <i>in vitro</i> or
CC	<i>in vivo</i> transfection and expression. They can thus be used for the
CC	delivery and expression of a therapeutic, immunological or immunogenic
CC	molecule (e.g., an antigen) and may also be used for eliciting an
CC	immunological response in a host organism. They are therefore useful in
CC	genetic vaccine compositions and for gene therapy, particularly where the
CC	use of retroviral vectors is unsafe or undesirable. Additionally, the
CC	retrotransposons may be used to generate transgenic animals, to detect
CC	the presence of <i>Candida</i> in a sample, to detect and disrupt genes, and to
CC	assign functions to nucleotide sequences. The present sequence represents
CC	the <i>C. albicans</i> pCal retrotransposon. (Updated on 15-SEP-2003 to
CC	standardise OS field)
XX	Sequence 6426 BP; 2228 A; 1032 C; 1258 G; 1908 T; 0 U; 0 Other;
SQ	
	Query Match 100.0%; Score 6424.4; DB 3; Length 6426;
	Best Local Similarity 100.0%; Pred. No. 0;
	Matches 6425; Conservative 0; Mismatches 1; Indels 0; Gaps 0
QY	1 TGTTGGTTGTCACATATTGTGTGCAGAACTGATCAATGAAGAAGTGTTATTATGA 50
DB	1 TGTTGGTTGTGCACATTTTGTGTGCAGAACTGATCAAGAAAAGATCGTTATTATGA 50
QY	61 GAATGAAAAATTTTTCATCAACAATCAGTGTGATGACGAACCTAATATATTGTGTACT 120
DB	61 GAATGAAAAATTTTTCATCAACAATCAGTGTGATGACGAACCTAATATTGTGTACT 120
QY	121 ATATAAAGAGGATGAATAATCCAAACATCCGGAATATCAACGATGAGAAGGAGAGATT 180
DB	121 ATATAAAGAGGATGAATAATCCAAACATCCGGAATATCAACGATGAGAAGGAGAGATT 180
QY	181 TCATATATATCTGTGTAATATACTGTGTTAATTCATATACAACTTAGACGTCT 240
DB	181 TCATATATATCTGTGTAATATACTGTGTTAATTCATATACAACTTAGACGTCT 240
QY	241 ACAAGCTCAATCTCAGATGAAGAACTTATATCCATCGATTAAGATGAGATAGTAT 300
DB	241 ACAAGCTCAATCTCAGATGAAGAACTTATATCCATCGATTAAGATGAGATAGTAT 300
QY	301 AACCATTTGTGCCCAAATTAGCGTTGATATAATTCACTGCTCAGATTGTATTATTGATT 360
DB	301 AACCATTTGTGCCCAAATTAGCGTTGATATAATTCACTGCTCAGATTGTATTATTGATT 360
QY	361 GATAGTTTGAAGTTGAAGTACGAATTTCAAGATAGTTTCCGCAAGAATGATGTA 420
DB	361 GATAGTTTGAAGTTGAAGTACGAATTTCAAGATAGTTTCCGCAAGAATGATGTA 420

```
QY 421 TAACGAGGAAAGTCATGAGAAAGTGGATCAAGCTAATGCTATTAGTAGTGATGA 480
DB 421 TAACGAGGAAAGTCATGAGAAAGTGGATCAAGCTAATGCTATTAGTAGTGATGA 480
QY 481 ACATATCAAGGCTAGATTCAATATGCTTTTCATAAATTTAAATGAAGCTTAAGTTGGC 540
DB 481 ACATATCAAGGCTAGATTCAATATGCTTTTCATAAATTTAAATGAAGCTTAAGTTGGC 540
QY 541 CGTCGATATCAGAAAGGCGGATTAATGGAATGAAGATTTAAATTTCCAGTTGC 600
DB 541 CGTCGATATCAGAAAGGCGGATTAATGGAATGAAGATTTAAATTTCCAGTTGC 600
QY 601 TTACCCCGATGTTTGGAAATTTTTCGCTGACTAATACTTAAAGATAATTCAGAGTTAA 660
DB 601 TTACCCCGATGTTTGGAAATTTTTCGCTGACTAATACTTAAAGATAATTCAGAGTTAA 660
QY 661 AAAGGTAGAAGGTATTTTATTACTGTTGGTGGTTTGAAGATGGTTTACAGTCCATTTT 720
DB 661 AAAGGTAGAAGGTATTTTATTACTGTTGGTGGTTTGAAGATGGTTTACAGTCCATTTT 720
QY 721 TGATAGGTTCAAGATTGATCATGATTTCTAAGCTACCAAGCACTTGCAAAAGAGCAAA 780
DB 721 TGATAGGTTCAAGATTGATCATGATTTCTAAGCTACCAAGCACTTGCAAAAGAGCAAA 780
QY 781 CTTAATCAAGGCTGTTATGATGCTGTTTACTAATCTTAAAGATTATACCATTTACTAGTA 840
DB 781 CTTAATCAAGGCTGTTATGATGCTGTTTACTAATCTTAAAGATTATACCATTTACTAGTA 840
QY 841 GATCTTGCTGAAGTTGTTTAAAGTTGAACATGATGATGTTGCTTAACTTCCATA 900
DB 841 GATCTTGCTGAAGTTGTTTAAAGTTGAACATGATGATGTTGCTTAACTTCCATA 900
QY 901 TTTGCTGAGGTGGAAGAGAAAAGTTGAGAAATCTTACAACTTCAAAAGTTTTCGA 960
DB 901 TTTGCTGAGGTGGAAGAGAAAAGTTGAGAAATCTTACAACTTCAAAAGTTTTCGA 960
QY 961 TGAGATGTCGCTGATCTTCCAAATCTGATAGGTCAGATCTTGAATCTTCAATCATGTGAA 1020
DB 961 TGAGATGTCGCTGATCTTCCAAATCTGATAGGTCAGATCTTGAATCTTCAATCATGTGAA 1020
QY 1021 GAATCAGAGGCTTTAAGTTGTTTGAATATTCATGCTCATATCTAATCTAAGTGGAT 1080
DB 1021 GAATCAGAGGCTTTAAGTTGTTTGAATATTCATGCTCATATCTAATCTAAGTGGAT 1080
QY 1081 TCAAGCTGACATGATATCATCTGACTCCCAAGTTGCTCTACCAATGCTGAAGAAATGTC 1140
DB 1081 TCAAGCTGACATGATATCATCTGACTCCCAAGTTGCTCTACCAATGCTGAAGAAATGTC 1140
QY 1141 TGATATCTGATTTATGCTAGATTTGTTGATGCTTCAAGCAATGCTCAAGTGGAT 1200
DB 1141 TGATATCTGATTTATGCTAGATTTGTTGATGCTTCAAGCAATGCTCAAGTGGAT 1200
QY 1201 TATTTGTTAGTTTACACGACCCAGAGAAAAGAAAGAAACCAAGAGGAACTCACTGGA 1260
DB 1201 TATTTGTTAGTTTACACGACCCAGAGAAAAGAAAGAAACCAAGAGGAACTCACTGGA 1260
QY 1261 ACAATCTCAAAAGAAAGAAAGCTGAATCAAGAAAGAAAGAAATTAAGAAACATCCAAATCGA 1320
DB 1261 ACAATCTCAAAAGAAAGAAAGCTGAATCAAGAAAGAAAGAAATTAAGAAACATCCAAATCGA 1320
QY 1321 TAAAGATAAAGGTGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGTCACTGGAATGAAAAACAG 1380
DB 1321 TAAAGATAAAGGTGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGTCACTGGAATGAAAAACAG 1380
QY 1381 TGCTGCTCTTATTAATGTTGTAATGAATATACATTAATGACGAAAAACACAGTTTCCAGT 1440
DB 1381 TGCTGCTCTTATTAATGTTGTAATGAATATACATTAATGACGAAAAACACAGTTTCCAGT 1440
QY 1441 AGAAAAATCTCATCTCTTAAAGCTTCTTGAAGCTAATTTTAAAGGTTTAAAGGTT 1500
DB 1441 AGAAAAATCTCATCTCTTAAAGCTTCTTGAAGCTAATTTTAAAGGTTTAAAGGTT 1500
QY 1501 TAAAGATATCTAGTATGATGATCTGTCACCAATATCTGTTGAAACAAATTAAGATAT 1560
DB 1501 TAAAGATATCTAGTATGATGATCTGTCACCAATATCTGTTGAAACAAATTAAGATAT 1560

QY 1501 TAACAGATATCTAGTATGATGATCTGTCACCAATATCTGTTGAAACAAATTAAGATAT 1560
DB 1501 TAACAGATATCTAGTATGATGATCTGTCACCAATATCTGTTGAAACAAATTAAGATAT 1560
QY 1561 ATTGCTGAATGTTAAGAGCAGCAACATTAAGTTTCTGTTGCTGATGAGTGCTTACATTA 1620
DB 1561 ATTGCTGAATGTTAAGAGCAGCAACATTAAGTTTCTGTTGCTGATGAGTGCTTACATTA 1620
QY 1621 AGCAGATGTAATGTTGATCTAATTAATCAAGTCGATATGTCGATTAAGTAAAGAA 1680
DB 1621 AGCAGATGTAATGTTGATCTAATTAATCAAGTCGATATGTCGATTAAGTAAAGAA 1680
QY 1681 TACATGTAATTTCCAGAAAGTCTTTAATCTTGAAGTTGAACAAATTAAGAAAG 1740
DB 1681 TACATGTAATTTCCAGAAAGTCTTTAATCTTGAAGTTGAACAAATTAAGAAAG 1740
QY 1741 AGGATTTAATGCTTATTAATCAAGAAATGATGATGTTATTTAACCAAAATGTCCTC 1800
DB 1741 AGGATTTAATGCTTATTAATCAAGAAATGATGATGTTATTTAACCAAAATGTCCTC 1800
QY 1801 TACTATTAATGCTTCAAGAGAAATGCTGATCTTATATAGGTCCTCAATTCAGTA 1860
DB 1801 TACTATTAATGCTTCAAGAGAAATGCTGATCTTATATAGGTCCTCAATTCAGTA 1860
QY 1861 AGAATCTTTPAATGTTGATTTTATGATGATGTTTGGAGATATGTTGCTCAATGCTAA 1920
DB 1861 AGAATCTTTPAATGTTGATTTTATGATGATGTTTGGAGATATGTTGCTCAATGCTAA 1920
QY 1921 CCAAGATGACAAAGATTAATCAAGTATGATGAATGATGATCAATATCAAGAAATGATTA 1980
DB 1921 CCAAGATGACAAAGATTAATCAAGTATGATGAATGATGATCAATATCAAGAAATGATTA 1980
QY 1981 TAGTCTCGACATTAATTAATCTTTGACGAGGTTGATGTTTGAATGTTGAAATTTTC 2040
DB 1981 TAGTCTCGACATTAATTAATCTTTGACGAGGTTGATGTTTGAATGTTGAAATTTTC 2040
QY 2041 CCCATATGAGGTTGAACAAATGCTTACCAAGTGAAGTAAAGAAAGATTTATTAATTTCA 2100
DB 2041 CCCATATGAGGTTGAACAAATGCTTACCAAGTGAAGTAAAGAAAGATTTATTAATTTCA 2100
QY 2101 TTTGATGTCAAATCATATGTCATTTGAGAAATCTTGTGTTTACAAAATACAGGAGTCT 2160
DB 2101 TTTGATGTCAAATCATATGTCATTTGAGAAATCTTGTGTTTACAAAATACAGGAGTCT 2160
QY 2161 CGTACTTCAACTTCAAAAGAGGCTTCAAAAGATTTGCTGATGTTGAAGTATGCTATTT 2220
DB 2161 CGTACTTCAACTTCAAAAGAGGCTTCAAAAGATTTGCTGATGTTGAAGTATGCTATTT 2220
QY 2221 ATCGAATGCCCCAAGAGAAAGTCAATCATCATCTTCAGAAAGAAAGCCTCGAAGAGCA 2280
DB 2221 ATCGAATGCCCCAAGAGAAAGTCAATCATCATCTTCAGAAAGAAAGCCTCGAAGAGCA 2280
QY 2281 TGAAGATCTTCAATGATGATCTCTGCTGATTTAGTCCGAAATTAACAAGTGTATTT 2340
DB 2281 TGAAGATCTTCAATGATGATCTCTGCTGATTTAGTCCGAAATTAACAAGTGTATTT 2340
QY 2341 AACGTCGTTATGATGATGATCACTGAGTTTACATGGAAGAAATTAATCAAGAAACAGAA 2400
DB 2341 AACGTCGTTATGATGATGATCACTGAGTTTACATGGAAGAAATTAATCAAGAAACAGAA 2400
QY 2401 GGTAAAGATCTCTTAATTCACAGATTAAGATCTGGAATTAATCGTTTAAGATTAAGT 2460
DB 2401 GGTAAAGATCTCTTAATTCACAGATTAAGATCTGGAATTAATCGTTTAAGATTAAGT 2460
QY 2461 GGCATATCTCAGAGGATTAATGCTGAGTTCCAGAACCTTCTGATTTAGTCTAGTT 2520
DB 2461 GGCATATCTCAGAGGATTAATGCTGAGTTCCAGAACCTTCTGATTTAGTCTAGTT 2520
QY 2521 CGGTATTTGAGGAGACTTAAGCGGATATCTGCTGAGCTTAATGTTCTGCGAGGT 2580
DB 2521 CGGTATTTGAGGAGACTTAAGCGGATATCTGCTGAGCTTAATGTTCTGCGAGGT 2580
QY 2581 TGTTAATTAATGATTTTCAACAGATTTAAGGAATGCTTGAACACTTGTGCTCAAT 2640
DB 2581 TGTTAATTAATGATTTTCAACAGATTTAAGGAATGCTTGAACACTTGTGCTCAAT 2640
```

Db 2581 TGTAAATTAATGATTTTACACAGATTTACAGAGATGTTGTGACACTTGTCCACAAAT 2640  
Qy 2641 ACTCAAGTTGATTTATATATGATTTCAATATATCAATATGATGATCAACACACTCCAG 2700  
Db 2641 ACTCAAGTTGATTTATATATGATTTCAATATATCAATATGATGATCAACACACTCCAG 2700  
Qy 2701 TCGTTCACTCAAGGAGCAAAACCCCTTAATGTTGCTATTATCAATTAAGTAGGGAATTT 2760  
Db 2701 TCGTTCACTCAAGGAGCAAAACCCCTTAATGTTGCTATTATCAATTAAGTAGGGAATTT 2760  
Qy 2761 CTACCGGTTTCTTTTGGCATGATGTTGTGCTATCACTTTAGTAATGATTCGAAAAGA 2820  
Db 2761 CTACCGGTTTCTTTTGGCATGATGTTGTGCTATCACTTTAGTAATGATTCGAAAAGA 2820  
Qy 2821 CCGTTACGAGTTCATCACTTAAGAGAGCTCTTCATCGATCACTGGGCTGTGATGG 2880  
Db 2821 CCGTTACGAGTTCATCACTTAAGAGAGCTCTTCATCGATCACTGGGCTGTGATGG 2880  
Qy 2881 CTACGCTAGCGATTTGTTTATGTTATACGTTGCTTAAATAATGCGGTGTGATTTAT 2940  
Db 2881 CTACGCTAGCGATTTGTTTATGTTATACGTTGCTTAAATAATGCGGTGTGATTTAT 2940  
Qy 2941 CCTTAGCCCTTAATGCTGATATGCGAGCTATGAGGTTATTAATCTCTATCCAAA 3000  
Db 2941 CCTTAGCCCTTAATGCTGATATGCGAGCTATGAGGTTATTAATCTCTATCCAAA 3000  
Qy 3001 CTATTCACATACCTATATGATCACAATTTCTATGCTGATGATTCAGAGGAGCA 3060  
Db 3001 CTATTCACATACCTATATGATCACAATTTCTATGCTGATGATTCAGAGGAGCA 3060  
Qy 3061 ACTGGGCGCTCAGTACGAGGATACGCGAGACATATGAGGAAATGAAATGACATACAA 3120  
Db 3061 ACTGGGCGCTCAGTACGAGGATACGCGAGACATATGAGGAAATGAAATGACATACAA 3120  
Qy 3121 TGACGTGATGACATGCCCCAAGAGTCAATTTCACTTCAAGCCAGCATCGTTTACTTAA 3180  
Db 3121 TGACGTGATGACATGCCCCAAGAGTCAATTTCACTTCAAGCCAGCATCGTTTACTTAA 3180  
Qy 3181 TACGGGTACAGTTCTTAACGAATATGTTAAATGATGATTCAGATTAAGATTTACATTA 3240  
Db 3181 TACGGGTACAGTTCTTAACGAATATGTTAAATGATGATTCAGATTAAGATTTACATTA 3240  
Qy 3241 GAATCCGATGATTTTCTTAACCCCTTCACTTAAGTAAAGATCAACGATATGATTC 3300  
Db 3241 GAATCCGATGATTTTCTTAACCCCTTCACTTAAGTAAAGATCAACGATATGATTC 3300  
Qy 3301 CGAAGTAAATCGAGTGAATTCCTAAACCCAGTCTCCAGAGCTTAAACCTGGGAGTAA 3360  
Db 3301 CGAAGTAAATCGAGTGAATTCCTAAACCCAGTCTCCAGAGCTTAAACCTGGGAGTAA 3360  
Qy 3361 TCCGGTCTAAACCTCTCAACTGTGATCGAGAGCTTCACTAATGAGGAACTTAAGA 3420  
Db 3361 TCCGGTCTAAACCTCTCAACTGTGATCGAGAGCTTCACTAATGAGGAACTTAAGA 3420  
Qy 3421 GCTATTAACAAACACAAAGAGAGCCCTTCATCCAGGGAGGAGCACTAAAGCCT 3480  
Db 3421 GCTATTAACAAACACAAAGAGAGCCCTTCATCCAGGGAGGAGCACTAAAGCCT 3480  
Qy 3481 GGAATCTACTGCTCAGTTGATCTATCAACCAACCCAGAGCTGTCTCCCGTTGGA 3540  
Db 3481 GGAATCTACTGCTCAGTTGATCTATCAACCAACCCAGAGCTGTCTCCCGTTGGA 3540  
Qy 3541 GGAATCTACTGCTCAGTTGATCTATCAACCAACCCAGAGCTGTCTCCCGTTGGA 3600  
Db 3541 GGAATCTACTGCTCAGTTGATCTATCAACCAACCCAGAGCTGTCTCCCGTTGGA 3600  
Qy 3601 AGAATATGAGCATCTTCTGACTACCCGAAACTAGTCTGAAGATGAACGCAAAATCC 3660  
Db 3601 AGAATATGAGCATCTTCTGACTACCCGAAACTAGTCTGAAGATGAACGCAAAATCC 3660  
Qy 3661 CTGCTTACTGCTAATGAAGATCGGTAATGAAAAAATAGATAGGAGGAAATATTTTC 3720  
Db 3661 CTGCTTACTGCTAATGAAGATCGGTAATGAAAAAATAGATAGGAGGAAATATTTTC 3720

Qy 3721 ATTTCCGGGGGTGATGATGATTTCTGTGATCACTCAATGTTGAGCAATTAATGT 3780  
Db 3721 ATTTCCGGGGGTGATGATGATTTCTGTGATCACTCAATGTTGAGCAATTAATGT 3780  
Qy 3781 TGAAACAGAGATGCTGTTAACAGTCCCAATTCAGAGAGAGTTCCTCAAGGGAGAGAT 3840  
Db 3781 TGAAACAGAGATGCTGTTAACAGTCCCAATTCAGAGAGAGTTCCTCAAGGGAGAGAT 3840  
Qy 3841 ACTAATGAACAACTGATATAGTTGATGATGATGATGATGATGATGATGATGATGAT 3900  
Db 3841 ACTAATGAACAACTGATATAGTTGATGATGATGATGATGATGATGATGATGATGAT 3900  
Qy 3901 CTCTCTTAATTAATTAATGATGATCACTGAACTTGTCTCAAGCTCCGAAATGATAG 3960  
Db 3901 CTCTCTTAATTAATTAATGATGATCACTGAACTTGTCTCAAGCTCCGAAATGATAG 3960  
Qy 3961 CAATTCACAGATTCGACATTCGAAATGAAATATCACCAGTATTAATGAGA 4020  
Db 3961 CAATTCACAGATTCGACATTCGAAATGAAATATCACCAGTATTAATGAGA 4020  
Qy 4021 AAATACGAATTAATTCAAAAACACATTTGAAGTATCTGCTGATTAAGATTTGATGA 4080  
Db 4021 AAATACGAATTAATTCAAAAACACATTTGAAGTATCTGCTGATTAAGATTTGATGA 4080  
Qy 4081 ATTTGAACGATTAATGTTGATGATGATGATGATGATGATGATGATGATGATGAT 4140  
Db 4081 ATTTGAACGATTAATGTTGATGATGATGATGATGATGATGATGATGATGATGAT 4140  
Qy 4141 AGCTAATCCACTACAGATGAATTAATGATGATGATGATGATGATGATGATGATGAT 4200  
Db 4141 AGCTAATCCACTACAGATGAATTAATGATGATGATGATGATGATGATGATGATGAT 4200  
Qy 4201 TATAGCATGATGACAGCAAGAAATTAACATTTGAGAAAGATTAAGAAAT 4260  
Db 4201 TATAGCATGATGACAGCAAGAAATTAACATTTGAGAAAGATTAAGAAAT 4260  
Qy 4261 TCTGTGATTAACCTCACTGATGATGATGATGATGATGATGATGATGATGATGAT 4320  
Db 4261 TCTGTGATTAACCTCACTGATGATGATGATGATGATGATGATGATGATGATGAT 4320  
Qy 4321 TACTGTGAAACCAAGAGATGAAACCAAGAAAGAGAGTTCATCACTGGAAT 4380  
Db 4321 TACTGTGAAACCAAGAGATGAAACCAAGAAAGAGAGTTCATCACTGGAAT 4380  
Qy 4381 AAACAAAGATGCAATCACTGATGATGATGATGATGATGATGATGATGATGATGAT 4440  
Db 4381 AAACAAAGATGCAATCACTGATGATGATGATGATGATGATGATGATGATGATGAT 4440  
Qy 4441 CCCAAGAGTATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4500  
Db 4441 CCCAAGAGTATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4500  
Qy 4501 AAAGTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4560  
Db 4501 AAAGTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4560  
Qy 4561 AATTCACCGGTGGAAGCTTATCTATGAGTTCGATGATGATGATGATGATGATGAT 4620  
Db 4561 AATTCACCGGTGGAAGCTTATCTATGAGTTCGATGATGATGATGATGATGATGAT 4620  
Qy 4621 TCTCAAGGTGTTTCCGAAATCACTGTTGTTCTCAATGAGCAACAAAGGAA 4680  
Db 4621 TCTCAAGGTGTTTCCGAAATCACTGTTGTTCTCAATGAGCAACAAAGGAA 4680  
Qy 4681 ATTTGAATTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4740  
Db 4681 ATTTGAATTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4740  
Qy 4741 ATTTGAATTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4800  
Db 4741 ATTTGAATTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4800

QY 4801 GATCTAAATGCGCTCTATTACTCATTCGAATCCAAATTATGCTTCTCTCTAAATCAGT 4860  
 DB 4801 GATCTAAATGCGCTCTATTACTCATTCGAATCCAAATTATGCTTCTCTCTAAATCAGT 4860  
 QY 4861 ACCCTTGAAGAAAACCATTTGTTGTTATTTGAAAGCTTCTGTCTATGGGTTTAAACAGTC 4920  
 DB 4861 ACCCTTGAAGAAAACCATTTGTTGTTATTTGAAAGCTTCTGTCTATGGGTTTAAACAGTC 4920  
 QY 4921 GGGTTTGAATGCTATCACACTATCAAAAAGATTTGGAAGACATTTGTTTACTCAAGT 4980  
 DB 4921 GGGTTTGAATGCTATCACACTATCAAAAAGATTTGGAAGACATTTGTTTACTCAAGT 4980  
 QY 4981 TTTACCAATGATGGTTTATTTTCACTGAATATGAAAGGGAATAGAAATATATTTAGG 5040  
 DB 4981 TTTACCAATGATGGTTTATTTTCACTGAATATGAAAGGGAATAGAAATATATTTAGG 5040  
 QY 5041 TTTATATGTTGATGATATTTCTTATGTTGGAAGTTTCAAAAAGTTATTTGATTTTGT 5100  
 DB 5041 TTTATATGTTGATGATATTTCTTATGTTGGAAGTTTCAAAAAGTTATTTGATTTTGT 5100  
 QY 5101 GGATCAATGAGATGATCTTTTGAAGTTTGAAGTTTGGTGAATATCAAAATTTATCTTGG 5160  
 DB 5101 GGATCAATGAGATGATCTTTTGAAGTTTGAAGTTTGGTGAATATCAAAATTTATCTTGG 5160  
 QY 5161 TATGAAATTCGTAACCGAATCTGTTATATTTATCTCAAGAAAATTTCTCAAGAA 5220  
 DB 5161 TATGAAATTCGTAACCGAATCTGTTATATTTATCTCAAGAAAATTTCTCAAGAA 5220  
 QY 5221 ATTACTTAAGATTTTCAAACTTGATGACTCATATGGGAAAAATCATCTCTGATTCGAA 5280  
 DB 5221 ATTACTTAAGATTTTCAAACTTGATGACTCATATGGGAAAAATCATCTCTGATTCGAA 5280  
 QY 5281 TGACAAATATGAAAAAGTTGCAATATTCGTAATAAGCTTATCAGAGAAATGATTTTGA 5340  
 DB 5281 TGACAAATATGAAAAAGTTGCAATATTCGTAATAAGCTTATCAGAGAAATGATTTTGA 5340  
 QY 5341 AAAGGTTCCGATGAGACATTCGTTGACCTGATGCTAAAAAATCATACCAAGTGGTGT 5400  
 DB 5341 AAAGGTTCCGATGAGACATTCGTTGACCTGATGCTAAAAAATCATACCAAGTGGTGT 5400  
 QY 5401 TGGCGCTGTTTATGCGGCGCAACAACGCTCCAAATATTCGCTCGATGAAATTC 5460  
 DB 5401 TGGCGCTGTTTATGCGGCGCAACAACGCTCCAAATATTCGCTCGATGAAATTC 5460  
 QY 5461 GTTGGGTTCTAAATCTGCAAAATCCAAATGCTCATGATATGAGAAATTTATTTGCT 5520  
 DB 5461 GTTGGGTTCTAAATCTGCAAAATCCAAATGCTCATGATATGAGAAATTTATTTGCT 5520  
 QY 5521 TAGGATATTCAAAAATAGCATGGGATATCATATGAGTACAAAAGAAACAGATTTGAAAT 5580  
 DB 5521 TAGGATATTCAAAAATAGCATGGGATATCATATGAGTACAAAAGAAACAGATTTGAAAT 5580  
 QY 5581 ACCACCAAAATCATTTGTTATGCAATTTTCAAGTATCGTATTTGACACAGATTTGA 5640  
 DB 5581 ACCACCAAAATCATTTGTTATGCAATTTTCAAGTATCGTATTTGACACAGATTTGA 5640  
 QY 5641 TAGAAATCTATTTAGTGAACCTTGATTTATGTAATGGAATTTGGTGCATAGGGGAC 5700  
 DB 5641 TAGAAATCTATTTAGTGAACCTTGATTTATGTAATGGAATTTGGTGCATAGGGGAC 5700  
 QY 5701 CAAAAAACAACGGCTATGACAAAGCTGACGCTTGTGAATTTGGCTCTAAATTA 5760  
 DB 5701 CAAAAAACAACGGCTATGACAAAGCTGACGCTTGTGAATTTGGCTCTAAATTA 5760  
 QY 5761 TACAATGTTGAAGCTATGCAAAATTAACCATTTATATGATTTGGGTTTGAAGTAGG 5820  
 DB 5761 TACAATGTTGAAGCTATGCAAAATTAACCATTTATATGATTTGGGTTTGAAGTAGG 5820  
 QY 5821 TAAAGTATGATGATGATGCAAGCAACAGCTGTATTAAGTTTGAAGAAATATCTATTG 5880  
 DB 5821 TAAAGTATGATGATGATGCAAGCAACAGCTGTATTAAGTTTGAAGAAATATCTATTG 5880  
 QY 5881 TCACCACATGACCAATAGATATCTGCTATATAGTTTCTAGCCAAATGATCATATGATA 5940

DB 5881 TCACCACATGACCAATAGATATCTGCTATATAGTTTCTAGCCAAATGATCATATGATA 5940  
 QY 5941 AGTATTTTCAATATCCATATGTAAGACAAATGATAATTTAGCCGATTTGATGCTAAGT 6000  
 DB 5941 AGTATTTTCAATATCCATATGTAAGACAAATGATAATTTAGCCGATTTGATGCTAAGT 6000  
 QY 6001 TCTAAGTGTCTTAATTCAAATTCAAAGCATTTGTTGAGGGTATGATATAACGTTAGACTAGA 6060  
 DB 6001 TCTAAGTGTCTTAATTCAAATTCAAAGCATTTGTTGAGGGTATGATATAACGTTAGACTAGA 6060  
 QY 6061 AGATTAATCAACACTATTCAAAATGCAATTAACGCGAATATAGTGAATTTATCATTTACT 6120  
 DB 6061 AGATTAATCAACACTATTCAAAATGCAATTAACGCGAATATAGTGAATTTATCATTTACT 6120  
 QY 6121 ATTATGTAATGCTCAATCAGGGAGTGTGTTTGTGCACTATTTTGTGTCAGAAACTG 6180  
 DB 6121 ATTATGTAATGCTCAATCAGGGAGTGTGTTTGTGCACTATTTTGTGTCAGAAACTG 6180  
 QY 6181 ATCAATGAAAATGATGCTTATTTGGAATGGAATTTTTCATCATCATCAGCTGAT 6240  
 DB 6181 ATCAATGAAAATGATGCTTATTTGGAATGGAATTTTTCATCATCATCAGCTGAT 6240  
 QY 6241 GACAGAACTTAACTATTTGTATGATATTAATTAAGGGTATGAAATCCAAATCCAGAA 6300  
 DB 6241 GACAGAACTTAACTATTTGTATGATATTAATTAAGGGTATGAAATCCAAATCCAGAA 6300  
 QY 6301 TATCAACGAGATGGAAGGAGGAGTTTCAATATATCTTTGGAATTAATTAACCTGCTCT 6360  
 DB 6301 TATCAACGAGATGGAAGGAGGAGTTTCAATATATCTTTGGAATTAATTAACCTGCTCT 6360  
 QY 6361 AATTCACTATACCACTGAGCTGTAACGCTCAATTCAGGTAAGAAAGTTATATT 6420  
 DB 6361 AATTCACTATACCACTGAGCTGTAACGCTCAATTCAGGTAAGAAAGTTATATT 6420  
 QY 6421 CCATCA 6426  
 DB 6421 CCATCA 6426  
 RESULT 2  
 AAAS7936  
 ID AAAS7936 standard; DNA; 392 Bp.  
 AC AAAS7936;  
 XX  
 DT 15-SEP-2003 (revised)  
 DT 10-OCT-2000 (first entry)  
 DE Candida albicans strain p36 Tca2 retrotransposon 5' region.  
 XX  
 KW Retrotransposon; pCa1; Tca2; Ty1; copia; long terminal repeat; LTR;  
 KW 58S gene; group antigen; pol; polyprotein; pol; aspartate protease; integrase;  
 KW reverse transcriptase; RNaseH; pseudonuc; readthrough translation;  
 KW scop codon suppression; gene delivery; gene therapy vector;  
 KW genetic vaccine composition; immunogenic; transgenic animal; ds.  
 XX  
 OS Candida albicans; strain p36.  
 XX  
 PN W0200026397-A1.  
 PD 11-MAY-2000.  
 XX  
 PF 01-NOV-1999; 99WO-NZ000179.  
 XX  
 PR 30-OCT-1998; 98CA-02249046.  
 PR 30-OCT-1998; 98US-0106342P.  
 XX  
 PA (JANC) JANSSEN PHARM NV.  
 PI Luyten WHM, De Backer MD, Nelissen BJW, Pouliet RTM;  
 XX  
 DR WPI; 2000-365640/31.

XX Novel retrotransposon expression vectors useful for expressing an  
PT antigen, epitope or therapeutic agent, or detecting genes or the presence  
of *Candida* in a sample.

PS Example 9; Fig 9; 204pp; English.

XX The invention relates to novel retrotransposons from the yeast *Candida*  
CC albicans which have a copy number of 40-150, preferably 50-100 copies per  
CC genome. In particular, the invention relates to the novel *C. albicans*  
CC Ty1/copia retrotransposon pCal (AA57920), and to the integrated form of  
CC this retrotransposon, designated TCa2, and to the novel *C. albicans*  
CC retrotransposons 1-28. pCal was initially isolated from *C. albicans*  
CC HOG1042 and has a copy number of 50-100 copies per cell. It comprises  
CC identical 280 bp long terminal repeats (LTRs) and two open reading frames  
CC (ORFs). The first ORF encodes a gag (group antigen) protein, and the  
CC second ORF encodes a polypeptide (pol) consisting of an aspartate  
CC protease, integrase, reverse transcriptase (RT), and RNaseH. The gag and  
CC pol ORFs of pCal are in the same reading frame, separated only by a  
CC termination codon (TGA). Translation of the pol ORF occurs through the  
CC occasional readthrough suppression of the stop codon, which is mediated  
CC by the formation of a pseudoknot within the gag-pol mRNA. The  
CC retrotransposons of the invention can be used as vectors for in vitro or  
CC in vivo transformation and expression. They can thus be used for the  
CC delivery and expression of a therapeutic, immunological or immunogenic  
CC molecule (e.g., an antigen) and may also be used for eliciting an  
CC immunological response in a host organism. They are therefore useful in  
CC genetic vaccine compositions and for gene therapy, particularly where the  
CC use of retroviral vectors is unsafe or undesirable. Additionally, the  
CC retrotransposons may be used to generate transgenic animals, to detect  
CC the presence of *Candida* in a sample, to detect and disrupt genes, and to  
CC assign functions to nucleotide sequences. Sequences AA57923-457936  
CC represents TCa2 retrotransposon 5' regions from a variety of *C. albicans*  
CC strains. (Updated on 15-SEP-2003 to standardise OS field)

XX Sequence 392 BP; 136 A; 56 C; 72 G; 128 T; 0 U; 0 Other;

Query Match 6.1%; Score 392; DB 3; Length 392;  
Best Local Similarity 100.0%; Pred. No. 1.8e-69;  
Matches 392; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGTGGTTGGCACTATTTGTGCGAAGATGAAATGAGTATATATGA 60  
DB 1 TGTGGTTGGCACTATTTGTGCGAAGATGAAATGAGTATATATGA 60  
QY 61 GAATGAAATTTTCCATCACAATCAGTATGACAGAACTAACTATATGTAGT 120  
DB 61 GAATGAAATTTTCCATCACAATCAGTATGACAGAACTAACTATATGTAGT 120  
QY 121 ATAAATTAAGGGTATGAAATACCAACATCCAGAAATATCAAGAAATGAGGAGGAGTT 180  
DB 121 ATAAATTAAGGGTATGAAATACCAACATCCAGAAATATCAAGAAATGAGGAGGAGTT 180  
QY 181 TCAATATATATCTTGTGAAATATACTCGTTCTTAATTCATATACCAACTAGACGTGT 240  
DB 181 TCAATATATATCTTGTGAAATATACTCGTTCTTAATTCATATACCAACTAGACGTGT 240  
QY 241 ACACGCTCAATCTCAGGTAAAGAAATTTATTTTCATCAATTTGAAGTGAATGAT 300  
DB 241 ACACGCTCAATCTCAGGTAAAGAAATTTATTTTCATCAATTTGAAGTGAATGAT 300  
QY 301 AATCATTTGCGCCAAATTAAGGTGATTAATTCAGTCCAGATTTGATTTATTTATTT 360  
DB 301 AATCATTTGCGCCAAATTAAGGTGATTAATTCAGTCCAGATTTGATTTATTTATTT 360  
QY 361 GATAGTTTGAAGTTGAAGGTACAGAAATTTTC 392  
DB 361 GATAGTTTGAAGTTGAAGGTACAGAAATTTTC 392

RESULT 3  
AA57928  
ID AA57928 standard; DNA; 392 BP.

XX AA57928;  
XX 15-SEP-2003 (revised)  
DT 10-OCT-2000 (first entry)  
XX

DE *Candida albicans* strain ATC-2 TCa2 retrotransposon 5' region.

XX Retrotransposon; pCal; TCa2; Ty1; copia; long terminal repeat; LTR;  
XX gag gene; group antigen; polypeptide; pol; aspartate protease; integrase;  
XX reverse transcriptase; RNaseH; pseudoknot; readthrough translation;  
XX stop codon suppression; gene delivery; gene therapy vector;  
XX genetic vaccine composition; immunogenic; transgenic animal; ds.

XX *Candida albicans*; strain ATC-2.

XX WO200026397-A1.

XX 11-MAY-2000.

XX 01-NOV-1999; 99WO-N2000179.

XX 30-OCT-1998; 98CA-02249046.

XX 30-OCT-1998; 98US-0106342P.

XX (JANC) JANSSEN PHARM NV.

XX Luyten WHML, De Backer MD, Nelissen BM, Poulier RTM;  
XX WPI; 2000-365640/31.

PT Novel retrotransposon expression vectors useful for expressing an  
PT antigen, epitope or therapeutic agent, or detecting genes or the presence  
of *Candida* in a sample.

PS Example 9; Fig 9; 204pp; English.

XX The invention relates to novel retrotransposons from the yeast *Candida*  
CC albicans which have a copy number of 40-150, preferably 50-100 copies per  
CC genome. In particular, the invention relates to the novel *C. albicans*  
CC Ty1/copia retrotransposon pCal (AA57920), and to the integrated form of  
CC this retrotransposon, designated TCa2, and to the novel *C. albicans*  
CC retrotransposons 1-28. pCal was initially isolated from *C. albicans*  
CC HOG1042 and has a copy number of 50-100 copies per cell. It comprises  
CC identical 280 bp long terminal repeats (LTRs) and two open reading frames  
CC (ORFs). The first ORF encodes a gag (group antigen) protein, and the  
CC second ORF encodes a polypeptide (pol) consisting of an aspartate  
CC protease, integrase, reverse transcriptase (RT) and RNaseH. The gag and  
CC pol ORFs of pCal are in the same reading frame, separated only by a  
CC termination codon (TGA). Translation of the pol ORF occurs through the  
CC occasional readthrough suppression of the stop codon, which is mediated  
CC by the formation of a pseudoknot within the gag-pol mRNA. The  
CC retrotransposons of the invention can be used as vectors for in vitro or  
CC in vivo transformation and expression. They can thus be used for the  
CC delivery and expression of a therapeutic, immunological or immunogenic  
CC molecule (e.g., an antigen) and may also be used for eliciting an  
CC immunological response in a host organism. They are therefore useful in  
CC genetic vaccine compositions and for gene therapy, particularly where the  
CC use of retroviral vectors is unsafe or undesirable. Additionally, the  
CC retrotransposons may be used to generate transgenic animals, to detect  
CC the presence of *Candida* in a sample, to detect and disrupt genes, and to  
CC assign functions to nucleotide sequences. Sequences AA57923-457936  
CC represents TCa2 retrotransposon 5' regions from a variety of *C. albicans*  
CC strains. (Updated on 15-SEP-2003 to standardise OS field)

XX Sequence 392 BP; 136 A; 56 C; 72 G; 128 T; 0 U; 0 Other;

Query Match 6.1%; Score 388.8; DB 3; Length 392;  
Best Local Similarity 99.5%; Pred. No. 7.9e-69;  
Matches 390; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGTGGTTGGCACTATTTGTGCGAAGATGAAATGAGTATATATGA 60  
|||

Db 1 TGTGTTTGGACATATTTTGTGTCAGAACTGATCATGATAAATGATGTTATATGA 60  
 QY 61 GAATGGAAAAATTTTTCATCAGATCAGTGATGACGAACTAATATTTGTGTA 120  
 Db 61 GAATGGAAAAATTTTTCATCAGATCAGTGATGACGAACTAATATTTGTGTA 120  
 QY 121 ATAATTAAGGATGATGAAATACCAACATCCGAAATATCAAGAGATAGAGAGAGATT 180  
 Db 121 ATAATTAAGGATGATGAAATACCAACATCCGAAATATCAAGAGATAGAGAGAGATT 180  
 QY 181 TCAATATATATCTTGTGATATATACTGTTCTTAATTCAGTATACACAACTAGACGTT 240  
 Db 181 TCAATATATATCTTGTGATATATACTGTTCTTAATTCAGTATACACAACTAGACGTT 240  
 QY 241 ACAGGCTCAATCTCAGTAAAGAAAGTTTATATTCATCAGATTAAGAGTGAATGAT 300  
 Db 241 ACAGGCTCAATCTCAGTAAAGAAAGTTTATATTCATCAGATTAAGAGTGAATGAT 300  
 QY 301 AATCATTTTGTCCCAATTAAGGCTGTATATAATTCAGTCTCAGATTTGTATTTGATT 360  
 Db 301 AATCATTTTGTCCCAATTAAGGCTGTATATAATTCAGTCTCAGATTTGTATTTGATT 360  
 QY 361 GATAGTTTCGAGATTGGAAGGTACAGAAATTTTC 392  
 Db 361 GATAGTTTCGAGATTGGAAGGTACAGAAATTTTC 392  
 RESULT 4  
 ID AAA57931 standard; DNA; 392 BP.  
 AC AAA57931;  
 XX 15-SEP-2003 (revised)  
 DT 10-OCT-2000 (first entry)  
 XX  
 DE Candida albicans strain Fl6-1 Tca2 retrotransposon 5' region.  
 XX  
 KW Retrotransposon; pCal; Tca2; Ty1; copia; long terminal repeat; LTR;  
 KW gag gene; group antigen; polyprotein; pol; aspartate protease; integrase;  
 KW reverse transcriptase; RNaseH; pseudoknot; readthrough translation;  
 KW stop codon suppression; gene delivery; gene therapy vector;  
 KW genetic vaccine composition; immunogenic; transgenic animal; ds.  
 XX  
 OS Candida albicans; strain Fl6-1.  
 XX  
 PN WO20026397-A1.  
 XX  
 PD 11-MAY-2000.  
 XX  
 PF 01-NOV-1999; 99WO-NZ000179.  
 XX  
 PR 30-OCT-1998; 98CA-02249046.  
 PR 30-OCT-1998; 98US-0106342P.  
 XX  
 PA (JUNC) JANSSEN PHARM NV.  
 XX  
 PI Luyten WHML, De Backer MD, Nelissen BJM, Poultier RTM;  
 DR WPI; 2000-365640/31.  
 PT Novel retrotransposon expression vectors useful for expressing an  
 PT antigen, epitope or therapeutic agent, or detecting genes or the presence  
 PT of Candida in a sample.  
 XX  
 PS Example 9; Fig 9; 204pp; English.  
 XX  
 CC The invention relates to novel retrotransposons from the yeast Candida  
 CC albicans which have a copy number of 40-150, preferably 50-100 copies per  
 CC genome. In particular, the invention relates to the novel C. albicans  
 CC Ty1/copia retrotransposon pCal (AAA57920), and to the integrated form of  
 CC this retrotransposon, designated Tca2, and to the novel C. albicans  
 CC retrotransposons 1-28. pCal was initially isolated from C. albicans

CC hOG1042 and has a copy number of 50-100 copies per cell. It comprises  
 CC identical 280 bp long terminal repeats (LTRs) and two open reading frames  
 CC (ORFs). The first ORF encodes a gag (group antigen) protein, and the  
 CC second ORF encodes a polyprotein (pol) consisting of an aspartate  
 CC protease, integrase, reverse transcriptase (RT) and RNaseH. The gag and  
 CC pol ORFs of pCal are in the same reading frame, separated only by a  
 CC termination codon (TGA). Translation of the pol ORF occurs through the  
 CC occasional readthrough suppression of the stop codon, which is mediated  
 CC by the formation of a pseudoknot within the gag-pol mRNA. The  
 CC retrotransposons of the invention can be used as vectors for in vitro or  
 CC in vivo transformation and expression. They can thus be used for the  
 CC delivery and expression of a therapeutic, immunological or immunogenic  
 CC molecule (e.g., an antigen) and may also be used for eliciting an  
 CC immunological response in a host organism. They are therefore useful in  
 CC the use of retroviral vectors is unsafe or undesirable. Additionally, the  
 CC retrotransposons may be used to generate transgenic animals, to detect  
 CC the presence of Candida in a sample, to detect and disrupt genes, and to  
 CC assign functions to nucleotide sequences. Sequences AAA57923-457936  
 CC represent Tca2 retrotransposon 5' regions from a variety of C. albicans  
 CC strains. (Updated on 15-SEP-2003 to standardise OS field)  
 CC  
 SQ Sequence 392 BP; 136 A; 56 C; 72 G; 128 T; 0 U; 0 Other;  
 Query Match 6.1%; Score 388.8; DB 3; Length 392;  
 Best Local Similarity 99.5%; Pred. No. 7.9e-69;  
 Matches 390; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 TGTGTTTGGACATATTTTGTGTCAGAACTGATCATGATAAATGATGTTATATGA 60  
 Db 1 TGTGTTTGGACATATTTTGTGTCAGAACTGATCATGATAAATGATGTTATATGA 60  
 QY 61 GAATGGAAAAATTTTTCATCAGATCAGTGATGACGAACTAATATTTGTGTA 120  
 Db 61 GAATGGAAAAATTTTTCATCAGATCAGTGATGACGAACTAATATTTGTGTA 120  
 QY 121 ATAATTAAGGATGATGAAATACCAACATCCGAAATATCAAGAGATAGAGAGAGATT 180  
 Db 121 ATAATTAAGGATGATGAAATACCAACATCCGAAATATCAAGAGATAGAGAGAGATT 180  
 QY 181 TCAATATATATCTTGTGATATATACTGTTCTTAATTCAGTATACACAACTAGACGTT 240  
 Db 181 TCAATATATATCTTGTGATATATACTGTTCTTAATTCAGTATACACAACTAGACGTT 240  
 QY 241 ACAGGCTCAATCTCAGTAAAGAAAGTTTATATTCATCAGATTAAGAGTGAATGAT 300  
 Db 241 ACAGGCTCAATCTCAGTAAAGAAAGTTTATATTCATCAGATTAAGAGTGAATGAT 300  
 QY 301 AATCATTTTGTCCCAATTAAGGCTGTATATAATTCAGTCTCAGATTTGTATTTGATT 360  
 Db 301 AATCATTTTGTCCCAATTAAGGCTGTATATAATTCAGTCTCAGATTTGTATTTGATT 360  
 QY 361 GATAGTTTCGAGATTGGAAGGTACAGAAATTTTC 392  
 Db 361 GATAGTTTCGAGATTGGAAGGTACAGAAATTTTC 392  
 RESULT 5  
 ID AAA57923 standard; DNA; 392 BP.  
 AC AAA57923;  
 XX 15-SEP-2003 (revised)  
 DT 10-OCT-2000 (first entry)  
 XX  
 DE Candida albicans strain SGV-1 Tca2 retrotransposon 5' region.  
 XX  
 KW Retrotransposon; pCal; Tca2; Ty1; copia; long terminal repeat; LTR;  
 KW gag gene; group antigen; polyprotein; pol; aspartate protease; integrase;  
 KW reverse transcriptase; RNaseH; pseudoknot; readthrough translation;  
 KW stop codon suppression; gene delivery; gene therapy vector;  
 KW genetic vaccine composition; immunogenic; transgenic animal; ds.



```

XX OS Candida albicans; strain SGY-1.
XX PN WO200026397-A1.
XX PD 11-MAY-2000.
XX PF 01-NOV-1999; 99WO-NZ000179.
XX PR 30-OCT-1998; 98CA-02249046.
XX PR 30-OCT-1998; 98US-0106342P.
XX PA (JANC ) JANSSEN PHARM NV.
XX PI Luyten WML, De Backer MD, Nelissen BOM, Poultier RTM;
XX DR WPI; 2000-365640/31.
XX PT Novel retrotransposon expression vectors useful for expressing an
XX PT antigen, epitope or therapeutic agent, or detecting genes or the presence
XX PT of Candida in a sample.
XX PS Example 9; Fig 9; 204pp; English.
XX CC The invention relates to novel retrotransposons from the yeast Candida
XX CC albicans which have a copy number of 40-150, preferably 50-100 copies per
XX CC genome. In particular, the invention relates to the novel C. albicans
XX CC Ty1/copia retrotransposon pCal (AA57920), and to the integrated form of
XX CC this retrotransposon, designated TCa2, and to the novel C. albicans
XX CC retrotransposons 1-28. pCal was initially isolated from C. albicans
XX CC HOG1042 and has a copy number of 50-100 copies per cell. It comprises
XX CC identical 280 bp long terminal repeats (LTRs) and two open reading frames
XX CC (ORFs). The first ORF encodes a gag (group antigen) protein, and the
XX CC second ORF encodes a polypeptide (pol) consisting of an aspartate
XX CC protease, integrase, reverse transcriptase (RT) and RNaseH. The gag and
XX CC pol ORFs of pCal are in the same reading frame, separated only by a
XX CC termination codon (TGA). Translation of the pol ORF occurs through the
XX CC occasional readthrough suppression of the stop codon, which is mediated
XX CC by the formation of a pseudoknot within the gag-pol mRNA. The
XX CC retrotransposons of the invention can be used as vectors for in vitro or
XX CC in vivo transformation and expression. They can thus be used for the
XX CC delivery and expression of a therapeutic, immunological or immunogenic
XX CC molecule (e.g., an antigen) and may also be used for eliciting an
XX CC immunological response in a host organism. They are therefore useful in
XX CC genetic vaccine compositions and for gene therapy, particularly where the
XX CC use of retroviral vectors is unsafe or undesirable. Additionally, the
XX CC retrotransposons may be used to generate transgenic animals, to detect
XX CC the presence of Candida in a sample, to detect and disrupt genes, and to
XX CC assign functions to nucleotide sequences. Sequences AA57923-AA57936
XX CC represent TCa2 retrotransposon 5' regions from a variety of C. albicans
XX CC strains. (Updated on 15-SEP-2003 to standardise OS field)
XX SQ Sequence 392 BP; 136 A; 56 C; 72 G; 128 T; 0 U; 0 Other;
XX
XX Query Match 6.1%; Score 388.8; DB 3; Length 392;
XX Best Local Similarity 99.5%; Pred. No. 7.9e-69;
XX Matches 390; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX 1 TGTGTTGTCATATTTGTGTGTCAGAAATGATCAATGATGATGATTTATTTGA 60
XX DB 1 TGTGTTGTCATATTTGTGTGTCAGAAATGATCAATGATGATGATTTATTTGA 60
XX
XX QY 61 GAATGAAAATTTTTCATCACAATCAGTGATGACAGAACTAACTATTTGTGAGT 120
XX DB 61 GAATGAAAATTTTTCATCACAATCAGTGATGACAGAACTAACTATTTGTGAGT 120
XX
XX QY 121 ATAATTAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 180
XX DB 121 ATAATTAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 180
XX
XX QY 181 TCAATATATATCTTGTGATATATTAATCTGTTCTTATTTACTATATACAACTTGA 240
XX DB 181 TCAATATATATCTTGTGATATATTAATCTGTTCTTATTTACTATATACAACTTGA 240

```

```

QY 241 ACAGCTTATCTCAGGTAAAGAAAGTTTATTCATTCAGATTAGAGTGCATAGTAT 300
DB 241 ACAGCTTATCTCAGGTAAAGAAAGTTTATTCATTCAGATTAGAGTGCATAGTAT 300
QY 301 AATCATTTGTCGCCAAATTTAGCTTGTATTAATTCAGTCCATGATTTGATTTGAT 360
DB 301 AATCATTTGTCGCCAAATTTAGCTTGTATTAATTCAGTCCATGATTTGATTTGAT 360
QY 361 GATAGTTTCGAAAGTTTGAGGTACAGAAATTC 392
DB 361 GATAGTTTCGAAAGTTTGAGGTACAGAAATTC 392
XX
XX RESULT 5
XX AA57932
XX ID AA57932 standard; DNA; 392 BP.
XX
XX AC AA57932;
XX XX
XX DT 15-SEP-2003 (revised)
XX DT 10-OCT-2000 (first entry)
XX
XX DE Candida albicans strain Fl6-2 TCa2 retrotransposon 5' region.
XX
XX XX Retrotransposon; pCal; TCa2; Ty1; copia; long terminal repeat; LTR;
XX XX gag gene; group antigen; polypeptide; pol; aspartate protease; integrase;
XX XX reverse transcriptase; RNaseH; pseudoknot; readthrough translation;
XX XX stop codon suppression; gene delivery; gene therapy vector;
XX XX genetic vaccine composition; immunogenic; transgenic animal; de.
XX
XX OS Candida albicans; strain Fl6-2.
XX
XX PN WO200026397-A1.
XX PD 11-MAY-2000.
XX PF 01-NOV-1999; 99WO-NZ000179.
XX PR 30-OCT-1998; 98CA-02249046.
XX PR 30-OCT-1998; 98US-0106342P.
XX PA (JANC ) JANSSEN PHARM NV.
XX PI Luyten WML, De Backer MD, Nelissen BOM, Poultier RTM;
XX DR WPI; 2000-365640/31.
XX
XX PS Example 9; Fig 9; 204pp; English.
XX
XX CC The invention relates to novel retrotransposons from the yeast Candida
XX CC albicans which have a copy number of 40-150, preferably 50-100 copies per
XX CC genome. In particular, the invention relates to the novel C. albicans
XX CC Ty1/copia retrotransposon pCal (AA57920), and to the integrated form of
XX CC this retrotransposon, designated TCa2, and to the novel C. albicans
XX CC retrotransposons 1-28. pCal was initially isolated from C. albicans
XX CC HOG1042 and has a copy number of 50-100 copies per cell. It comprises
XX CC identical 280 bp long terminal repeats (LTRs) and two open reading frames
XX CC (ORFs). The first ORF encodes a gag (group antigen) protein, and the
XX CC second ORF encodes a polypeptide (pol) consisting of an aspartate
XX CC protease, integrase, reverse transcriptase (RT) and RNaseH. The gag and
XX CC pol ORFs of pCal are in the same reading frame, separated only by a
XX CC termination codon (TGA). Translation of the pol ORF occurs through the
XX CC occasional readthrough suppression of the stop codon, which is mediated
XX CC by the formation of a pseudoknot within the gag-pol mRNA. The
XX CC retrotransposons of the invention can be used as vectors for in vitro or
XX CC in vivo transformation and expression. They can thus be used for the
XX CC delivery and expression of a therapeutic, immunological or immunogenic
XX CC molecule (e.g., an antigen) and may also be used for eliciting an

```



CC immunological response in a host organism. They are therefore useful in  
 CC genetic vaccine compositions and for gene therapy, particularly where the  
 CC use of retroviral vectors is unsafe or undesirable. Additionally, the  
 CC retrotransposons may be used to generate transgenic animals, to detect  
 CC the presence of *Candida* in a sample, to detect and disrupt genes, and to  
 CC assign functions to nucleotide sequences. Sequences AA57923-AS7936  
 CC represents Tca2 retrotransposon 5' regions from a variety of *C. albicans*  
 CC strains. (Updated on 15-SEP-2003 to standardise OS field)

XX Sequence 392 BP; 136 A; 56 C; 72 G; 128 T; 0 U; 0 Other;

Query Match 6.1%; Score 388.8; DB 3; Length 392;  
 Best Local Similarity 99.5%; Pred. No. 7.9e-69;  
 Matches 390; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGTGGTTGGTGCATATTTTGTGCAGAACTGATCAATGAAATGATGTTTATGA 60  
 DB 1 TGTGGTTGGTGCATATTTTGTGCAGAACTGATCAATGAAATGATGTTTATGA 60  
 QY 61 GAATGAAAAATTTTCCATCACAATCAGTGATGACAGAACTAACTATATTTGTAGT 120  
 DB 61 GAATGAAAAATTTTCCATCACAATCAGTGATGACAGAACTAACTATATTTGTAGT 120  
 QY 121 ATAAATTAAGGGTATTAATTCACATCCAGAAATTAAGAGATTAAGAGAGATT 180  
 DB 121 ATAAATTAAGGGTATTAATTCACATCCAGAAATTAAGAGATTAAGAGAGATT 180  
 QY 181 TCAATATATATCTTGATATATACTCGTTCTAATTCAGTATATACCACTAAGCGTGT 240  
 DB 181 TCAATATATATCTTGATATATACTCGTTCTAATTCAGTATATACCACTAAGCGTGT 240  
 QY 241 ACACGCTCAATCTCGGTAAAGAAATTTATTTCCATCAATGAAATGATGAT 300  
 DB 241 ACACGCTCAATCTCGGTAAAGAAATTTATTTCCATCAATGAAATGATGAT 300  
 QY 301 AATCATTTGCTCCCAATTAAGGCTGTATTAATTCAGTCTCAGATTTGATTTGATT 360  
 DB 301 AATCATTTGCTCCCAATTAAGGCTGTATTAATTCAGTCTCAGATTTGATTTGATT 360  
 QY 361 GATAGTTTGGAGTTTGAAGGTACAGAAATTTTC 392  
 DB 361 GATAGTTTGGAGTTTGAAGGTACAGAAATTTTC 392

## RESULT 7

AA57927  
 ID AA57927 standard; DNA; 392 BP.

AC AA57927;  
 XX  
 DT 15-SEP-2003 (revised)  
 DT 10-OCT-2000 (first entry)

DE *Candida albicans* strain ATC-1 Tca2 retrotransposon 5' region.  
 XX  
 XX Retrotransposon; pCal; Tca2; Ty1; copia; long terminal repeat; LTR;  
 KW gag gene; group antigen; polyprotein; pol; aspartate protease; integrase;  
 KW reverse transcriptase; RNaseH; pseudoknot; readthrough translation;  
 KW stop codon suppression; gene delivery; gene therapy vector;  
 KW genetic vaccine composition; immunogenic; transgenic animal; ds.

XX *Candida albicans*; strain ATC-1.

XX MO200026397-AL.

XX 11-MAY-2000.

XX 01-NOV-1999; 99WO-NZ000179.

XX 30-OCT-1998; 98CA-02249046.  
 XX 30-OCT-1998; 98US-0106342P.

XX (JANSEN PHARM NV.

XX Luyten WML, De Backer MD, Nelissen BJM, Poulter RTM;  
 XX WPI; 2000-365640/31.  
 DR  
 PT Novel retrotransposon expression vectors useful for expressing an  
 PT antigen, epitope or therapeutic agent, or detecting genes or the presence  
 PT of *Candida* in a sample.

XX Example 9; Fig 9; 204pp; English.

XX The invention relates to novel retrotransposons from the yeast *Candida*  
 CC *albicans* which have a copy number of 40-150, preferably 50-100 copies per  
 CC genome. In particular, the invention relates to the novel *C. albicans*  
 CC Ty1/copia retrotransposon pCal (AA57920), and to the integrated form of  
 CC this retrotransposon, designated Tca2, and to the novel *C. albicans*  
 CC retrotransposons 1-28. pCal was initially isolated from *C. albicans*  
 CC H0G1042 and has a copy number of 50-100 copies per cell. It comprises  
 CC identical 280 bp long terminal repeats (LTRs) and two open reading frames  
 CC (ORFs). The first ORF encodes a gag (group antigen) protein, and the  
 CC second ORF encodes a polyprotein (pol) consisting of an aspartate  
 CC protease, integrase, reverse transcriptase (RT) and RNaseH. The gag and  
 CC pol ORFs of pCal are in the same reading frame, separated only by a  
 CC termination codon (TGA). Translation of the pol ORF occurs through the  
 CC occasional readthrough suppression of the stop codon, which is mediated  
 CC by the formation of a pseudoknot within the gag-pol mRNA. The  
 CC retrotransposons of the invention can be used as vectors for in vitro or  
 CC in vivo transformation and expression. They can thus be used for the  
 CC delivery and expression of a therapeutic, immunological or immunogenic  
 CC molecule (e.g., an antigen) and may also be used for eliciting an  
 CC immunological response in a host organism. They are therefore useful in  
 CC genetic vaccine compositions and for gene therapy, particularly where the  
 CC use of retroviral vectors is unsafe or undesirable. Additionally, the  
 CC retrotransposons may be used to generate transgenic animals, to detect  
 CC the presence of *Candida* in a sample, to detect and disrupt genes, and to  
 CC assign functions to nucleotide sequences. Sequences AA57923-AS7936  
 CC represents Tca2 retrotransposon 5' regions from a variety of *C. albicans*  
 CC strains. (Updated on 15-SEP-2003 to standardise OS field)

XX Sequence 392 BP; 136 A; 56 C; 72 G; 128 T; 0 U; 0 Other;

Query Match 6.1%; Score 388.8; DB 3; Length 392;  
 Best Local Similarity 99.5%; Pred. No. 7.9e-69;  
 Matches 390; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGTGGTTGGTGCATATTTTGTGCAGAACTGATCAATGAAATGATGTTTATGA 60  
 DB 1 TGTGGTTGGTGCATATTTTGTGCAGAACTGATCAATGAAATGATGTTTATGA 60  
 QY 61 GAATGAAAAATTTTCCATCACAATCAGTGATGACAGAACTAACTATATTTGTAGT 120  
 DB 61 GAATGAAAAATTTTCCATCACAATCAGTGATGACAGAACTAACTATATTTGTAGT 120  
 QY 121 ATAAATTAAGGGTATTAATTCACATCCAGAAATTAAGAGATTAAGAGAGATT 180  
 DB 121 ATAAATTAAGGGTATTAATTCACATCCAGAAATTAAGAGATTAAGAGAGATT 180  
 QY 181 TCAATATATATCTTGATATATACTCGTTCTAATTCAGTATATACCACTAAGCGTGT 240  
 DB 181 TCAATATATATCTTGATATATACTCGTTCTAATTCAGTATATACCACTAAGCGTGT 240  
 QY 241 ACACGCTCAATCTCGGTAAAGAAATTTATTTCCATCAATGAAATGATGAT 300  
 DB 241 ACACGCTCAATCTCGGTAAAGAAATTTATTTCCATCAATGAAATGATGAT 300  
 QY 301 AATCATTTGCTCCCAATTAAGGCTGTATTAATTCAGTCTCAGATTTGATTTGATT 360  
 DB 301 AATCATTTGCTCCCAATTAAGGCTGTATTAATTCAGTCTCAGATTTGATTTGATT 360  
 QY 361 GATAGTTTGGAGTTTGAAGGTACAGAAATTTTC 392  
 DB 361 GATAGTTTGGAGTTTGAAGGTACAGAAATTTTC 392





QY 181 TCAATATATATCTTGTGATATATATCTGCTGTAATTCACATACACAACTAGAGTGT 240  
 Db 181 TCAATATATATCTTGTGATATATATCTGCTGTAATTCACATACAACTAGAGTGT 240  
 QY 241 ACACGCTCAATCTCAGGTAAAGAAAGTTTATTCATTCAGATTAGAGTCGATAGTAT 300  
 Db 241 ACACGCTCAATCTCAGGTAAAGAAAGTTTATTCATTCAGATTAGAGTCGATAGTAT 300  
 QY 301 AATCATTTTCGTCCTCAATTTAGCGTTGTATTAATTCAGTCTCTCAGATTGTATATGATT 360  
 Db 301 AATCATTTTCGTCCTCAATTTAGCGTTGTATTAATTCAGTCTCTCAGATTGTATATGATT 360  
 QY 361 GATAGTTTCGAGATTGAGGTACAGAAATTTTC 392  
 Db 361 GATAGTTTCGAGATTGAGGTACAGAAATTTTC 392

RESULT 11  
 AA57930  
 ID AA57930 standard; DNA; 392 BP.

XX AA57930;  
 XX 15-SEP-2003 (revised)  
 DT 10-OCT-2000 (first entry)

DE Candida albicans strain SA4-2 Tca2 retrotransposon 5' region.

XX Retrotransposon; pCal; Tca2; Ty1; copia; long terminal repeat; LTR;  
 KM gag gene; group antigen; polyprotein; pol; aspartate protease; integrase;  
 KM reverse transcriptase; RNaseH; pseudoknot; readthrough translation;  
 KM stop codon suppression; gene delivery; gene therapy vector;  
 KM genetic vaccine composition; immunogenic; transgenic animal; ds.

XX Candida albicans; strain SA4-2.

XX WO200026397-A1.

XX 11-MAY-2000.

XX 01-NOV-1999; 99WO-NZ000179.

XX 30-OCT-1998; 98CA-02249046.

XX 30-OCT-1998; 98US-0106342P.

XX (JANC) JANSSEN PHARM NV.

XX Luyken WHML, De Backer MD, Nelissen BJM, Poultier RTM;

XX WPI; 2000-365640/31.

PT Novel retrotransposon expression vectors useful for expressing an  
 PT antigen, epitope or therapeutic agent, or detecting genes or the presence  
 PT of Candida in a sample.

XX Example 9; Fig 9; 204pp; English.

CC The invention relates to novel retrotransposons from the yeast Candida  
 CC albicans which have a copy number of 40-150, preferably 50-100 copies per  
 CC genome. In particular, the invention relates to the novel C. albicans  
 CC Ty1/copia retrotransposon pCal (AA57920), and to the integrated form of  
 CC this retrotransposon, designated Tca2, and to the novel C. albicans  
 CC retrotransposons 1-28. pCal was initially isolated from C. albicans  
 CC HO31042 and has a copy number of 50-100 copies per cell. It comprises  
 CC identical 280 bp long terminal repeats (LTRs) and two open reading frames  
 CC (ORFs). The first ORF encodes a gag (group antigen) protein, and the  
 CC second ORF encodes a polyprotein (pol) consisting of an aspartate  
 CC protease, integrase, reverse transcriptase (RT) and RNaseH. The gag and  
 CC pol ORFs of pCal are in the same reading frame, separated only by a  
 CC termination codon (TGA). Translation of the pol ORF occurs through the  
 CC occasional readthrough suppression of the stop codon, which is mediated  
 CC by the formation of a pseudoknot within the gag-pol mRNA. The

CC retrotransposons of the invention can be used as vectors for in vitro or  
 CC in vivo transformation and expression. They can thus be used for the  
 CC delivery and expression of a therapeutic, immunological or immunogenic  
 CC molecule (e.g., an antigen) and may also be used for eliciting an  
 CC immunological response in a host organism. They are therefore useful in  
 CC genetic vaccine compositions and for gene therapy, particularly where the  
 CC use of retroviral vectors is unsafe or undesirable. Additionally, the  
 CC retrotransposons may be used to generate transgenic animals, to detect  
 CC the presence of Candida in a sample, to detect and disrupt genes, and to  
 CC assign functions to nucleotide sequences. Sequences AA57923-A57936  
 CC represent Tca2 retrotransposon 5' regions from a variety of C. albicans  
 CC strains. (Updated on 15-SEP-2003 to standardise OS field)

XX Sequence 392 BP; 135 A; 58 C; 72 G; 127 T; 0 U; 0 Other;

XX Query Match 6.0%; Score 385.6; DB 3; Length 392;

XX Best Local Similarity 99.0%; Pred. No. 3.5e-68;

XX Matches 388; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGTGTGTTTGTCACATATTTTGTGTGCAAACTGATGAAATGATGTTATTTATGCA 60  
 Db 1 TGTGTGTTTGTCACATATTTTGTGTGCAAACTGATGAAATGATGTTATTTATGCA 60  
 QY 61 GAATGAAATTTTTCATCATCAGATGATGACAGAACTAACTATTTGTAGT 120  
 Db 61 GAATGAAATTTTTCATCATCAGATGATGACAGAACTAACTATTTGTAGT 120  
 QY 121 ATAAATTAAGGTATGCAATATCCAACTCCGAAATATCAAGATATGAAGGAGGAGTT 180  
 Db 121 ATAAATTAAGGTATGCAATATCCAACTCCGAAATATCAAGATATGAAGGAGGAGTT 180  
 QY 181 TCAATATATATCTTGTGATATATATCTGCTGTAATTCACATACAACTAGAGTGT 240  
 Db 181 TCAATATATATCTTGTGATATATATCTGCTGTAATTCACATACAACTAGAGTGT 240  
 QY 241 ACACGCTCAATCTCAGGTAAAGAAAGTTTATTCATTCAGATTAGAGTCGATAGTAT 300  
 Db 241 ACACGCTCAATCTCAGGTAAAGAAAGTTTATTCATTCAGATTAGAGTCGATAGTAT 300  
 QY 301 AATCATTTTCGTCCTCAATTTAGCGTTGTATTAATTCAGTCTCTCAGATTGTATATGATT 360  
 Db 301 AATCATTTTCGTCCTCAATTTAGCGTTGTATTAATTCAGTCTCTCAGATTGTATATGATT 360  
 QY 361 GATAGTTTCGAGATTGAGGTACAGAAATTTTC 392  
 Db 361 GATAGTTTCGAGATTGAGGTACAGAAATTTTC 392

RESULT 12  
 AA57934  
 ID AA57934 standard; DNA; 392 BP.

XX AA57934;

XX 15-SEP-2003 (revised)

XX 10-OCT-2000 (first entry)

DE Candida albicans strain 759-2 Tca2 retrotransposon 5' region.

XX Retrotransposon; pCal; Tca2; Ty1; copia; long terminal repeat; LTR;  
 KM gag gene; group antigen; polyprotein; pol; aspartate protease; integrase;  
 KM reverse transcriptase; RNaseH; pseudoknot; readthrough translation;  
 KM stop codon suppression; gene delivery; gene therapy vector;  
 KM genetic vaccine composition; immunogenic; transgenic animal; ds.

XX Candida albicans; strain 759-2.

XX WO200026397-A1.

XX 11-MAY-2000.

XX 01-NOV-1999; 99WO-NZ000179.

PR 30-OCT-1998; 98CA-02249046.  
 PR 30-OCT-1998; 98US-0106342P.  
 XX (JANC ) JANSSEN PHARM NV.  
 PA  
 XX Luyten WHML, De Backer MD, Nelissen BJM, Poultier RTM;  
 PI MPI: 2000-365640/31.  
 XX  
 PT Novel retrotransposon expression vectors useful for expressing an  
 PT antigen, epitope or therapeutic agent, or detecting genes or the presence  
 PT of Candida in a sample.  
 XX  
 PS Example 9; Fig 9; 204pp; English.  
 XX  
 CC The invention relates to novel retrotransposons from the yeast *Candida*  
 CC albicans which have a copy number of 40-150, preferably 50-100 copies per  
 CC genome. In particular, the invention relates to the novel *C. albicans*  
 CC Ty1/copia retrotransposon pcal (AA57920), and to the integrated form of  
 CC this retrotransposon, designated Tca2, and to the novel *C. albicans*  
 CC retrotransposons 1-28. pcal was initially isolated from *C. albicans*  
 CC HOG1042 and has a copy number of 50-100 copies per cell. It comprises  
 CC identical 280 bp long terminal repeats (LTRs) and two open reading frames  
 CC (ORFs). The first ORF encodes a gag (group antigen) protein, and the  
 CC second ORF encodes a polypeptide (pol) consisting of an aspartate  
 CC protease, integrase, reverse transcriptase (RT) and RNaseH. The gag and  
 CC pol ORFs of pcal are in the same reading frame, separated only by a  
 CC termination codon (TGA). Translation of the pol ORF occurs through the  
 CC occasional readthrough suppression of the stop codon, which is mediated  
 CC by the formation of a pseudoknot within the gag-pol mRNA. The  
 CC retrotransposons of the invention can be used as vectors for in vitro or  
 CC in vivo transformation and expression. They can thus be used for the  
 CC delivery and expression of a therapeutic, immunological or immunogenic  
 CC molecule (e.g., an antigen) and may also be used for eliciting an  
 CC immunological response in a host organism. They are therefore useful in  
 CC genetic vaccine compositions and for gene therapy, particularly where the  
 CC use of retroviral vectors is unsafe or undesirable. Additionally, the  
 CC retrotransposons may be used to generate transgenic animals, to detect  
 CC the presence of *Candida* in a sample, to detect and disrupt genes, and to  
 CC assign functions to nucleotide sequences. Sequences AA57923-A57936  
 CC represents Tca2 retrotransposon 5' regions from a variety of *C. albicans*  
 CC strains. (Updated on 15-SEP-2003 to standardise OS field)  
 CC  
 SO Sequence 392 BP; 138 A; 59 C; 71 G; 124 T; 0 U; 0 Other;  
 Query Match 6.0%; Score 384; DB 3; Length 392;  
 Best Local Similarity 98.7%; Pred. No. 7.4e-68;  
 Matches 387; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 TGTGGTTTGACATATTTTGTGACAAATGATCATATAAATGATGTTTATGCA 60  
 DB 1 TGTGGTTTGACATATTTTGTGACAAATGATCATATAAATGATGTTTATGCA 60  
 QY 61 GAATGAAATTTTCCATCACAATCGATGACAGAACTAACTATTTGTAGT 120  
 DB 61 GAATGAAATTTTCCATCACAATCGATGACAGAACTAACTATTTGTAGT 120  
 QY 121 ATTAATTAAGGTATGAATACCAACATCCCAATATCAACGAGATGAGGAGATT 180  
 DB 121 ATTAATTAAGGTATGAATACCAACATCCCAATATCAACGAGATGAGGAGATT 180  
 QY 121 ATTAATTAAGGTATGAATACCAACATCCCAATATCAACGAGATGAGGAGATT 180  
 DB 121 ATTAATTAAGGTATGAATACCAACATCCCAATATCAACGAGATGAGGAGATT 180  
 QY 181 TCAATATATCTTGTAATATACTGTTTCAATCACTATACCACTAAGCTGT 240  
 DB 181 TCAATATATCTTGTAATATACTGTTTCAATCACTATACCACTAAGCTGT 240  
 QY 241 ACACGCTCAATCTCAGTAAAGAAAGTTATATTCATCAGATTGAGAGTATGAT 300  
 DB 241 ACACGCTCAATCTCAGTAAAGAAAGTTATATTCATCAGATTGAGAGTATGAT 300  
 QY 301 AATCATTTGTCGCAATTAAGCGTTGTAATTAATCAGTCCCAATTTGTTATGAT 360  
 DB 301 AATCATTTGTCGCAATTAAGCGTTGTAATTAATCAGTCCCAATTTGTTATGAT 360

QY 361 GATAGTTGCAAGTTTGAAGCTACAGAAATTC 392  
 DB 361 GATAGTTGCAAGTTTGAAGCTACAGAAATTC 392  
 RESULT 13  
 ID AA57935 standard; DNA; 392 BP.  
 XX  
 AC AA57935;  
 XX  
 DT 15-SEP-2003 (revised)  
 DT 10-OCT-2000 (first entry)  
 XX  
 DE *Candida albicans* strain p30 Tca2 retrotransposon 5' region.  
 XX  
 KW Retrotransposon; pcal; Tca2; Ty1; copia; long terminal repeat; LTR;  
 KW gag gene; group antigen; polyprotein; pol; aspartate protease; integrase;  
 KW reverse transcriptase; RNaseH; pseudoknot; readthrough translation;  
 KW stop codon suppression; gene delivery; gene therapy vector;  
 KW genetic vaccine composition; immunogenic; transgenic animal; ds.  
 XX  
 OS *Candida albicans*; strain p30.  
 XX  
 PN WO200026397-A1.  
 XX  
 PD 11-MAY-2000.  
 XX  
 PF 01-NOV-1999; 99WC-NZ000179.  
 XX  
 PR 30-OCT-1998; 98CA-02249046.  
 PR 30-OCT-1998; 98US-0106342P.  
 XX  
 PA (JANC ) JANSSEN PHARM NV.  
 XX  
 PI Luyten WHML, De Backer MD, Nelissen BJM, Poultier RTM;  
 XX MPI: 2000-365640/31.  
 XX  
 DR Novel retrotransposon expression vectors useful for expressing an  
 PT antigen, epitope or therapeutic agent, or detecting genes or the presence  
 PT of *Candida* in a sample.  
 XX  
 PS Example 9; Fig 9; 204pp; English.  
 XX  
 CC The invention relates to novel retrotransposons from the yeast *Candida*  
 CC albicans which have a copy number of 40-150, preferably 50-100 copies per  
 CC genome. In particular, the invention relates to the novel *C. albicans*  
 CC Ty1/copia retrotransposon pcal (AA57920), and to the integrated form of  
 CC this retrotransposon, designated Tca2, and to the novel *C. albicans*  
 CC retrotransposons 1-28. pcal was initially isolated from *C. albicans*  
 CC HOG1042 and has a copy number of 50-100 copies per cell. It comprises  
 CC identical 280 bp long terminal repeats (LTRs) and two open reading frames  
 CC (ORFs). The first ORF encodes a gag (group antigen) protein, and the  
 CC second ORF encodes a polypeptide (pol) consisting of an aspartate  
 CC protease, integrase, reverse transcriptase (RT) and RNaseH. The gag and  
 CC pol ORFs of pcal are in the same reading frame, separated only by a  
 CC termination codon (TGA). Translation of the pol ORF occurs through the  
 CC occasional readthrough suppression of the stop codon, which is mediated  
 CC by the formation of a pseudoknot within the gag-pol mRNA. The  
 CC retrotransposons of the invention can be used as vectors for in vitro or  
 CC in vivo transformation and expression. They can thus be used for the  
 CC delivery and expression of a therapeutic, immunological or immunogenic  
 CC molecule (e.g., an antigen) and may also be used for eliciting an  
 CC immunological response in a host organism. They are therefore useful in  
 CC genetic vaccine compositions and for gene therapy, particularly where the  
 CC use of retroviral vectors is unsafe or undesirable. Additionally, the  
 CC retrotransposons may be used to generate transgenic animals, to detect  
 CC the presence of *Candida* in a sample, to detect and disrupt genes, and to  
 CC assign functions to nucleotide sequences. Sequences AA57923-A57936  
 CC represents Tca2 retrotransposon 5' regions from a variety of *C. albicans*  
 CC strains. (Updated on 15-SEP-2003 to standardise OS field)  
 CC  
 XX

SQ Sequence 392 BP; 139 A; 55 C; 71 G; 127 T; 0 U; 0 Other;  
Query Match 6.0%; Score 384; DB 3; Length 392;  
Best Local Similarity 98.7%; Pred. No. 7,4e-68;  
Matches 387; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```
OY 1 TGTGGTTTGGCACTATTTTGTGTGCAAAATCGATCAATGAAAATGATGTTATATGA 60
DB 1 TGTGGTTTGGCACTATTTTGTGTGCAAAATCGATCAATGAAAATGATGTTATATGA 60
OY 61 GAATGAAAATTTTTCATCATCATCAGGTGATGACAGAACTAACTATATTTGTAGT 120
DB 61 GAATGAAAATTTTTCATCATCATCAGGTGATGACAGAACTAACTATATTTGTAGT 120
OY 121 ATAAATAAGGGTATGAAATACCAACATCCGAAATATCAACGATGAAAGGAGAT 180
DB 121 ATAAATAAGGGTATGAAATACCAACATCCGAAATATCAACGATGAAAGGAGAT 180
OY 181 TCATATATATCTTGTGAAATATATCTGTTCAATCTCTATACCACTAGACGTGT 240
DB 181 TCATATATATCTTGTGAAATATATCTGTTCAATCTCTATACCACTAGACGTGT 240
OY 241 ACACGCTCAATCTCAGGTAAAGAAAGTTATATTCATCATCATGATTAGAGTATGAT 300
DB 241 ACACGCTCAATCTCAGGTAAAGAAAGTTATATTCATCATCATGATTAGAGTATGAT 300
OY 301 AATCATTTTCGTCCTCAATTAAGCGTTGTATTAATTCAGTCTCGATTGTATATGAT 360
DB 301 AATCATTTTCGTCCTCAATTAAGCGTTGTATTAATTCAGTCTCGATTGTATATGAT 360
OY 361 GATAGTTTCGAAGTTTGAAGTACAGAAATTC 392
DB 361 GATAGTTTCGAAGTTTGAAGTACAGAAATTC 392
```

RESULT 14  
ID AA57924  
AA57924 standard; DNA; 392 BP.

```
AC AA57924;
XX
XX 15-SEP-2003 (revised)
DT 10-OCT-2000 (first entry)
XX
XX Candida albicans strain SGY-2 TCa2 retrotransposon 5' region.
DE
XX Retrotransposon; pCal; TCa2; Tyl; copias; long terminal repeat; LTR;
KW gag gene; group antigen; polyprotein; pol; aspartate protease; integrase;
KW reverse transcriptase; RNaseH; pseudoknot; readthrough translation;
KW stop codon suppression; gene delivery; gene therapy vector;
KW genetic vaccine composition; immunogenic; transgenic animal; ds.
XX
XX Candida albicans; strain SGY-2.
OS
XX WO2000026397-A1.
XX
XX 11-MAY-2000.
XX
XX 01-NOV-1999; 99WO-NZ000179.
XX
XX 30-OCT-1998; 98CA-02249046.
XX
XX 30-OCT-1998; 98US-0106342P.
XX
XX (JANC ) JANSSEN PHARM NV.
XX
XX Luyten WHML, De Backer MD, Nelissen BJM, Poultier RTM;
XX
XX WPI; 2000-365640/31.
XX
XX Novel retrotransposon expression vectors useful for expressing an
XX antigen, epitope or therapeutic agent, or detecting genes or the presence
XX of Candida in a sample.
```

PS Example 9; Fig 9; 204pp; English.

XX The invention relates to novel retrotransposons from the yeast *Candida*  
CC albicans which have a copy number of 40-150, preferably 50-100 copies per  
CC genome. In particular, the invention relates to the novel *C. albicans*  
CC Tyl/copia retrotransposon pCal (AA57920), and to the integrated form of  
CC this retrotransposon, designated TCa2, and to the novel *C. albicans*  
CC retrotransposons 1-28. pCal was initially isolated from *C. albicans*  
CC HOG1042 and has a copy number of 50-100 copies per cell. It comprises  
CC identical 280 bp long terminal repeats (LTRs) and two open reading frames  
CC (ORFs). The first ORF encodes a gag (group antigen) protein, and the  
CC second ORF encodes a polypeptide (pol) consisting of an aspartate  
CC protease, integrase, reverse transcriptase (RT) and RNaseH. The gag and  
CC pol ORFs of pCal are in the same reading frame, separated only by a  
CC termination codon (TGA). Translation of the stop codon, which is mediated  
CC occasional readthrough suppression of the stop codon, results in the  
CC formation of a pseudoknot within the gag-pol mRNA. The gag-pol ORF  
CC retrotransposons of the invention can be used as vectors for in vitro or  
CC in vivo transformation and expression. They can thus be used for the  
CC delivery and expression of a therapeutic, immunological or immunogenic  
CC molecule (e.g., an antigen) and may also be used for eliciting an  
CC immunological response in a host organism. They are therefore useful in  
CC genetic vaccine compositions and for gene therapy, particularly where the  
CC use of retroviral vectors is unsafe or undesirable. Additionally, the  
CC retrotransposons may be used to generate transgenic animals, to detect  
CC the presence of *Candida* in a sample, to detect and disrupt genes, and to  
CC assign functions to nucleotide sequences. Sequences AA57923-AA57936  
CC represent TCa2 retrotransposon 5' regions from a variety of *C. albicans*  
CC strains. (Updated on 15-SEP-2003 to standardise OS field)

XX  
SQ Sequence 392 BP; 135 A; 57 C; 74 G; 126 T; 0 U; 0 Other;  
Query Match 6.0%; Score 382.4; DB 3; Length 392;  
Best Local Similarity 98.5%; Pred. No. 1.6e-67;  
Matches 386; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```
OY 1 TGTGGTTTGGCACTATTTTGTGTGCAAAATCGATCAATGAAAATGATGTTATATGA 60
DB 1 TGTGGTTTGGCACTATTTTGTGTGCAAAATCGATCAATGAAAATGATGTTATATGA 60
OY 61 GAATGAAAATTTTTCATCATCATCAGGTGATGACAGAACTAACTATATTTGTAGT 120
DB 61 GAATGAAAATTTTTCATCATCATCAGGTGATGACAGAACTAACTATATTTGTAGT 120
OY 121 ATAAATAAGGGTATGAAATACCAACATCCGAAATATCAACGATGAAAGGAGAT 180
DB 121 ATAAATAAGGGTATGAAATACCAACATCCGAAATATCAACGATGAAAGGAGAT 180
OY 181 TCATATATATCTTGTGAAATATATCTGTTCAATCTCTATACCACTAGACGTGT 240
DB 181 TCATATATATCTTGTGAAATATATCTGTTCAATCTCTATACCACTAGACGTGT 240
OY 241 ACACGCTCAATCTCAGGTAAAGAAAGTTATATTCATCATCATGATTAGAGTATGAT 300
DB 241 ACACGCTCAATCTCAGGTAAAGAAAGTTATATTCATCATCATGATTAGAGTATGAT 300
OY 301 AATCATTTTCGTCCTCAATTAAGCGTTGTATTAATTCAGTCTCGATTGTATATGAT 360
DB 301 AATCATTTTCGTCCTCAATTAAGCGTTGTATTAATTCAGTCTCGATTGTATATGAT 360
OY 361 GATAGTTTCGAAGTTTGAAGTACAGAAATTC 392
DB 361 GATAGTTTCGAAGTTTGAAGTACAGAAATTC 392
```

RESULT 15  
ID AA57926  
AA57926 standard; DNA; 392 BP.

```
AC AA57926;
XX
XX 15-SEP-2003 (revised)
DT 10-OCT-2000 (first entry)
XX
```

XX DE Candida albicans strain SC5-2 Tca2 retrotransposon 5' region.  
XX XX  
XX KW Retrotransposon; pCal; Tca2; Tyl; copia; long terminal repeat; LTR;  
KW gag gene; group antigen; pol; aspartate protease; integrase;  
KW reverse transcriptase; RNaseH; pseudonot; readthrough translation;  
KW stop codon suppression; gene delivery; gene therapy vector;  
XX KW genetic vaccine composition; immunogenic; transgenic animal; ds.  
XX OS Candida albicans; strain SC5-2.  
XX XX  
XX PN WO20026397-A1.  
XX PD 11-MAY-2000.  
XX PF 01-NOV-1999; 99WO-NZ000179.  
XX PR 30-OCT-1998; 98CA-02249046.  
XX PR 30-OCT-1998; 98US-0106342P.  
XX XX  
XX PA (JANCO) JANSSEN PHARM NV.  
XX PI Luyten WHM, De Backer MD, Nelissen BJM, Poultier RTM;  
XX DR WPI; 2000-365640/31.  
XX XX  
XX PT Novel retrotransposon expression vectors useful for expressing an  
XX PT antigen, epitope or therapeutic agent, or detecting genes or the presence  
XX PT of Candida in a sample.  
XX PS  
XX PS Example 9, Fig 9; 204pp; English.  
XX XX  
XX CC The invention relates to novel retrotransposons from the yeast Candida  
XX CC albicans which have a copy number of 40-150, preferably 50-100 copies per  
XX CC genome. In particular, the invention relates to the novel C. albicans  
XX CC Tyl1/copia retrotransposon pCal (AA57920), and to the integrated form of  
XX CC this retrotransposon, designated Tca2, and to the novel C. albicans  
XX CC retrotransposons 1-28. pCal was initially isolated from C. albicans  
XX CC hOG1042 and has a copy number of 50-100 copies per cell. It comprises  
XX CC identical 280 bp long terminal repeats (LTRs) and two open reading frames  
XX CC (ORFs). The first ORF encodes a gag (group antigen) protein, and the  
XX CC second ORF encodes a polypeptide (pol) consisting of an aspartate  
XX CC protease, integrase, reverse transcriptase (RT) and RNaseH. The gag and  
XX CC pol ORFs of pCal are in the same reading frame, separated only by a  
XX CC termination codon (TGA). Translation of the pol ORF occurs through the  
XX CC occasional readthrough suppression of the stop codon, which is mediated  
XX CC by the formation of a pseudoknot within the gag-pol mRNA. The  
XX CC retrotransposons of the invention can be used as vectors for in vitro or  
XX CC in vivo transformation and expression. They can thus be used for the  
XX CC delivery and expression of a therapeutic, immunological or immunogenic  
XX CC molecule (e.g., an antigen) and may also be used for eliciting an  
XX CC immunological response in a host organism. They are therefore useful in  
XX CC genetic vaccine compositions and for gene therapy, particularly where the  
XX CC use of retroviral vectors is unsafe or undesirable. Additionally, the  
XX CC retrotransposons may be used to generate transgenic animals, to detect  
XX CC the presence of Candida in a sample, to detect and disrupt genes, and to  
XX CC assign functions to nucleotide sequences. Sequences AA57923-AS7936  
XX CC represents Tca2 retrotransposon 5' regions from a variety of C. albicans  
XX CC strains. (Updated on 15-SEP-2003 to standardise OS field)  
XX SQ  
SQ Sequence 392 BP; 137 A; 57 C; 71 G; 127 T; 0 U; 0 Other;  
Query Match 5.9%; Score 380.8; DB 3; Length 392;  
Best Local Similarity 98.2%; Pred. No. 3.3e-67;  
Matches 385; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
QY 1 TGTGGTTGTGCACTATTTGTGTCAGAACTGATCAATGAAAATGATGGTTATATGA 60  
DB 1 TGTGGTTGTGCACTATTTGTGTCAGAACTGATCAATGAAAATGATGGTTATATGA 60  
QY 61 GAATGAAAATTTTCCATCACAATCAAGTGAATGACAGAACTAACTATTTGTAGT 120  
DB 61 GAATGAAAATTTTCCATCACAATCAAGTGAATGACAGAACTAACTATTTGTAGT 120

QY 121 ATAAATAGGTAATGAATTAACCAATCCAGATATCAAGAGATAGAGGAGAGT 180  
DB 121 ATAAATAGGTAATGAATTAACCAATCCAGATATCAAGAGATAGAGGAGAGT 180  
QY 181 TCAATATATATCTTGTGAATTAATACCTGTTCTAATTCATACACAACTAGACGTGT 240  
DB 181 TCAATATATATCTTGTGAATTAATACCTGTTCTAATTCATACACAACTAGACGTGT 240  
QY 241 ACAAGCTCATCTCCAGGTAAAGAAAGTTAATTCATCAGATTGAAGTCAATGAT 300  
DB 241 ACAAGCTCATCTCCAGGTAAAGAAAGTTAATTCATCAGATTGAAGTCAATGAT 300  
QY 301 AATCATTTGTCGCCAATTAAGCGTTGTATTAATTCAGTCCCTCAGATTGTATATGAT 360  
DB 301 AATCATTTGTCGCCAATTAAGCGTTGTATTAATTCAGTCCCTCAGATTGTATATGAT 360  
QY 361 GATAGTTTGAAGTTGAAAGGTACAGAAATTTT 392  
DB 361 GATAGTTTGAAGTTGAAAGGTACAGAAATTTT 392

Search completed: April 11, 2004, 14:02:32  
Job time : 1504 secs

This Page Blank (uspto)



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 11, 2004, 13:03:47 ; Search time 289 Seconds

(without alignments)  
12339.496 Million cell updates/sec

Title: US-09-430-590E-3

Perfect score: 6426

Sequence: 1 tggctgttgcacattc.....agaagttatctccatca 6426

Scoring table: IDENTITY NUC

Searched: 682709 seqs, 27747546 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued Patents.NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfiles.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	65.6	1.0	7218	1	US-08-232-463-14
2	58.2	0.9	640681	4	US-09-790-988-1
3	58	0.9	240	1	US-08-628-417-6
4	57.2	0.9	2394	4	US-09-800-729-33
C 5	56.8	0.9	5219	4	US-10-204-708-52
6	56.4	0.9	2447	2	US-09-014-968-14
7	56	0.9	5361	3	US-08-973-462-2
8	56	0.9	6152	3	US-08-973-462-1
C 9	55.6	0.9	396	4	US-09-640-173-53
10	55.6	0.9	396	4	US-09-713-550-53
11	55.6	0.9	11485	4	US-09-410-464-9
12	55.4	0.9	2826	4	US-09-254-776B-50
13	55.2	0.9	1696	4	US-09-835-811-1
14	54.6	0.8	6660	2	US-08-771-602D-1
15	54.6	0.8	6660	3	US-09-233-446B-1
16	54.4	0.8	3275	4	US-09-370-838-151
17	54	0.8	674	4	US-09-620-405B-465
18	54	0.8	674	4	US-09-433-826B-465
19	54	0.8	674	4	US-09-604-287A-465
20	54	0.8	674	4	US-09-834-759-465
C 21	53.6	0.8	5152	4	US-10-204-708-47
22	53.2	0.8	1798	4	US-09-797-906-1
23	53.2	0.8	8820	2	US-08-446-855A-1
24	53.2	0.8	8920	3	US-09-150-741-1
25	53	0.8	830	4	US-08-956-171E-715
26	52.8	0.8	1784	4	US-09-601-198-21
27	52.6	0.8	371	4	US-09-621-976-16048

C 28	52.4	0.8	5340	4	US-09-627-123-21	Sequence 21, Appl
29	52.2	0.8	14066	4	US-09-601-198-56	Sequence 56, Appl
30	52	0.8	2223	1	US-08-257-073-4	Sequence 4, Appl
31	52	0.8	30549	4	US-09-134-001C-322	Sequence 322, App
C 32	51.8	0.8	1020	4	US-09-328-475C-43	Sequence 43, Appl
33	51.4	0.8	1447	4	US-09-443-041A-27	Sequence 27, Appl
C 34	51	0.8	8607	4	US-10-204-708-72	Sequence 72, Appl
35	50.8	0.8	1866	4	US-09-601-198-153	Sequence 153, App
36	50.8	0.8	3095	6	5231168-1	Patent No. 5231168
37	50.2	0.8	6671	1	US-08-280-443-1	Sequence 1, Appl
38	50.2	0.8	6671	1	US-08-457-459-1	Sequence 1, Appl
39	50.2	0.8	6671	1	US-08-555-678-1	Sequence 1, Appl
C 40	50.2	0.8	6671	5	PCT-US95-02275-1	Sequence 29, Appl
C 41	50	0.8	260	2	US-08-520-678A-29	Sequence 29, Appl
C 42	50	0.8	260	3	US-08-897-125-29	Sequence 10, Appl
43	50	0.8	441	4	US-09-601-537-10	Sequence 509, App
44	50	0.8	1281	4	US-09-134-001C-509	Sequence 9, Appl
45	50	0.8	4121	4	US-09-601-537-9	

## ALIGNMENTS

RESULT 1  
US-08-232-463-14/c  
Sequence 14, Application US/08232463  
Patent No. 5670367  
GENERAL INFORMATION:  
APPLICANT: DORNER, F.  
APPLICANT: SCHEIFLINGER, F.  
APPLICANT: FALKNER, F. G.  
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232,463  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/935,313  
FILING DATE:  
APPLICATION NUMBER: BP 91 114 300.6  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET INFORMATION: 30472/114 INCU  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)683-4109  
TELEX: 889149  
INFORMATION FOR SEQ. ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: pTZ9p-Fls  
US-08-232-463-14  
Query Match 1.0%; Score 65.6; DB 1; Length 7218;



D <sub>b</sub>	74	AAA	133
Oy	1305	AACATCCAAATTCAGTATACGATTAAAGGTGAAGAAGAAAAGAAAGAACTTCAC	136
D <sub>b</sub>	134	AAA	193
Oy	1365	TGGAAATCAAAACA	1378
D <sub>b</sub>	194	AAAAAAAAAAAAAAAAA	207

RESULT 4  
US-09-800-729-33  
; Sequence 33, Application US/09800729  
; Patent No. 6605392  
; GENERAL INFORMATION:  
; ADDITIONAL INFORMATION:

```

1 TITLE OR INVENTION: 32 Human secreted proteins
2
3 FILE REFERENCE: P2044p1
4
5 CURRENT APPLICATION NUMBER: US/09/800,729
6
7 PRIOR APPLICATION NUMBER: PCT/US00/26013
8
9 PRIOR FILING DATE: 2000-09-22
10
11 PRIOR APPLICATION NUMBER: 60/155,709
12
13 PRIOR FILING DATE: 1998-09-24
14
15 NUMBER OF SEQ ID NOS: 217
16
17 SOFTWARE: Patenting Ver. 2.0
18
19 SEQ ID NO 33
20
21 LENGTH: 2394
22
23 TYPE: DNA
24
25 ORGANISM: Homo sapiens
26
27 US-09-800-729-33

```

Query Match	0.9%	Score 57.2	DB 4	Length 2394
Best Local Similarity	52.0%	Pred. No. 0.0012		
Matches 128; Conservative	0;	Mismatches 118;	Indels 0;	Gaps 0;

Qy	1193	GAATGTGTGATCATCCGATATATCTGATGGTGTACATTCGACGACCAATATGAA	1197
Db	2148	GAACCTCTTAATGATGCTTCTTCTTTCTTTAAAAATTCAGAAATCCAAATAAGSA	2207
Qy	1193	CTTAATCTATTGTTAGTTTACGACACGACAGAAACCAAAAGGAAAAACGAGGAGAAC	1257
Db	2208	AAGGACGGTAACTTATATGATGAAAAAATAAAAAAAGGAAAAAAGGAAAAA	2267
Qy	1253	TCCTCGGACATCTCCAAAAAGAGAACTGAAATACAGAAAGGAAATAGAAACATCCA	1313
Db	2268	AA	2327
Qy	1313	AAATCGATTAACGATTAAGGTGAAAAAGAAAAAGAAAAAACCCTCACTGGAATGA	1377
Db	2328	AA	2387
Qy	1373	AAAAACA	1378
Db	2388	AAAAAA	2393

```

RESULT 5
US-10-204-708-52/C
; Sequence 52, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; TITLE OF INVENTION: By Assessing DNA Methylation
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/204,708
; CURRENT FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06

```

? PRIOR APPLICATION NUMBER: DE 10019058.8  
 ? PRIOR FILING DATE: 2000-04-06  
 ? PRIOR APPLICATION NUMBER: DE 10019173.8  
 ? PRIOR FILING DATE: 2000-04-07  
 ? PRIOR APPLICATION NUMBER: DE 10032529.7  
 ? PRIOR FILING DATE: 2000-06-30  
 ? PRIOR APPLICATION NUMBER: DE 10043826.1  
 ? PRIOR FILING DATE: 2000-09-01  
 ? NUMBER OF SEQ ID NOS: 98  
 ? SEQ ID NO 52  
 ? LENGTH: 5219  
 ? TYPE: DNA  
 ? ORGANISM: Artificial Sequence  
 ? FEATURE:  
 ? OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens  
 ? US-10-204-708-52

Query Match	0.9%;	Score 56.8;	DB 4;	Length 5219;
Best Local Similarity	50.0%;	Pred. No. 0.0019;		
Matches 142;	Conservative	0;	Mismatches 142;	Indels 0;
				Gaps 0;

[illegible]

RESULT 6  
 US-09-014-969-14  
 Sequence 14, Application US/09014369  
 Patent No. 5965397  
 GENERAL INFORMATION:  
 APPLICANT: Jacobs, Kenneth  
 APPLICANT: McCoy, John M.  
 APPLICANT: Lavallee, Edward R.  
 APPLICANT: Racine, Lisa A.  
 APPLICANT: Metberg, David  
 APPLICANT: Treacy, Maurice  
 APPLICANT: Spaulding, Vikki  
 APPLICANT: Agostino, Michael J.  
 TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES  
 TITLE OF INVENTION: ENCODING THEM  
 NUMBER OF SEQUENCES: 32  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Genetics Institute, Inc.  
 STREET: 87 CambridgePark Drive  
 CITY: Cambridge  
 STATE: MA  
 COUNTRY: U.S.A.  
 ZIP: 02140  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/014,969  
 FILING DATE:

## CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:  
NAME: Springer, Suzanne A.  
REGISTRATION NUMBER: 41,323  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 498-8284  
TELEFAX: (617) 876-5851  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2447 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-09-014-969-14

## Query Match

0.9%; Score 56.4; DB 2; Length 2447;  
Best Local Similarity 53.8%; Pred. No. 0.0018;  
Matches 114; Conservative 1; Mismatches 97; Indels 0; Gaps 0;

QY 1167 TTGACATTCACAGCAGCAATATGACTTAATCTTATTTAGTTTACGACGACGAGAG 1226  
DB 2203 TTGATTTTGTACTTTAATGACAAATTAACCTTTGGGAGAAAAA 2262  
QY 1227 AACCAAAAGCAAAACCGAGAGACTCCTGAGACATCTCAAAAAGAACTGAAAT 1286  
DB 2263 AA 2322  
QY 1287 CAAGAAAGAGAAATAGAAACATCCAAATGAGATTAAGTGAAGAAAGAAAG 1346  
DB 2323 AA 2382  
QY 1347 AAAAAAAAAAACTCTGCTGATGAAACAA 1378  
DB 2383 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2414

## RESULT 7

US-08-973-462-2  
Sequence 2, Application US/08973462B  
Patent No. 6191270

GENERAL INFORMATION:  
APPLICANT: DRUILHE, PIERRE  
APPLICANT: DAUBERSTES, PIERRE  
TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES  
FILE REFERENCE: 0660-0125-0 PCT  
CURRENT APPLICATION NUMBER: US/08/973,462B  
EARLIER FILING DATE: 1998-02-06  
EARLIER APPLICATION NUMBER: PCT/FR96/00894  
EARLIER FILING DATE: 1996-06-12  
EARLIER APPLICATION NUMBER: FR 95/07007  
NUMBER OF SEQ ID NOS: 29  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 2  
LENGTH: 5361  
TYPE: DNA  
ORGANISM: P. falciparum  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)..(5361)  
US-08-973-462-2

## Query Match

0.9%; Score 56; DB 3; Length 5361;  
Best Local Similarity 45.9%; Pred. No. 0.003;  
Matches 191; Conservative 0; Mismatches 225; Indels 0; Gaps 0;

QY 3869 ACTGTTGCTAAAGTATGAGATGAAAAATCTCTCTATTAATCAATGATGATCAT 3928  
DB 289 AATGCTGATTAAGCTAGAGAGAGTAAAGAAAAATCTTGACTTAATGAGAGAGGA 348  
QY 3929 ACTGAACTTGCTACAGACTCGGAAATGATGCAATTCAGACAGATCGCATTCATCG 3988

DB 349 AATACATTACTGAAAGTGTAGATGATTAATTAATTTAGAGAGACCGGAGATTAAG 408  
QY 3989 AAAATGAAATATGACACGATTAATGAGAAAAATCTGAAATATCAAAAAACACTT 4048  
DB 409 GAAATATCTTTAATAGTAATAGTAAGAACCAAAAAATATTTAGCAATTTATTA 468  
QY 4049 GAAATATCTTTGCTGATTAAGAGTTGGATTAATTTGAAAGCTATTAATTTGATTAAT 4108  
DB 469 AATATATTTGACAAATTCAGAAAAACAAAAAGTATCAGAAAAATGTCAGAGT 528  
QY 4109 GAGAAATGATTAATGACGATGACATGCTGAAGCTATCCATACCAAGATGAAATAT 4168  
DB 529 GATGACTTTTAAATGATTAATTAATGATGATTAATGAGAGAGTAAAGAAAT 588  
QY 4169 GATGTTGATGAAATGAGAGTTTGTATATATCATGATGACGACGACGACGACGAC 4228  
DB 589 ATTTGAGAGAGAGCAAGTAAATGACGATTTTAAATGTTTAAAGTAAAGTTTCAA 648  
QY 4229 TACGATTTGAGAGAGAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 4284  
DB 649 CAAGAACAAACACACATGTTGAGAAAAAGTTGAAGAAAGTGAAGAAAGTGA 704

## RESULT 8

US-08-973-462-1  
Sequence 1, Application US/08973462B  
Patent No. 6191270

GENERAL INFORMATION:  
APPLICANT: DRUILHE, PIERRE  
APPLICANT: DAUBERSTES, PIERRE  
TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES  
FILE REFERENCE: 0660-0125-0 PCT  
CURRENT APPLICATION NUMBER: US/08/973,462B  
EARLIER FILING DATE: 1998-02-06  
EARLIER APPLICATION NUMBER: PCT/FR96/00894  
EARLIER FILING DATE: 1996-06-12  
EARLIER APPLICATION NUMBER: FR 95/07007  
NUMBER OF SEQ ID NOS: 29  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1  
LENGTH: 6152  
TYPE: DNA  
ORGANISM: P. falciparum  
US-08-973-462-1

Query Match  
0.9%; Score 56; DB 3; Length 6152;  
Best Local Similarity 45.9%; Pred. No. 0.0031;  
Matches 191; Conservative 0; Mismatches 225; Indels 0; Gaps 0;

QY 3869 ACTGTTGCTAAAGTATGAGATGAAAAATCTCTCTATTAATCAATGATGATCAT 3928  
DB 533 AATGCTGATTAAGCTAGAGAGAGTAAAGAAAAATCTTGACTTAATGAGAGAGGA 592  
QY 3929 ACTGAACTTGCTACAGACTCGGAAATGATGCAATTCAGACAGATTCGATTCG 3988  
DB 593 AATACATTAACTGAAAGTGTAGATGATTAATTAATTAAGAGAGCCGAGATTAAG 652  
QY 3989 AAAATGAAATATGACACGATTAATGAGAAAAATCTGAAATATCAAAAAACACTT 4048  
DB 653 GAAATATCTTTAATAGTAATAGTAAGAACCAAAAAATATTTAGCAATTTATTA 712  
QY 4049 GAAATATCTTTGCTGATTAAGAGTTGGATTAATTTGAAAGCTATTAATTTGATTAAT 4108  
DB 713 AATATATTTGACAAATTCAGAAAAACAAAAAGTATCAGAAAAATGTCAGAGT 772  
QY 4109 GAGAAATGATTAATGACGATGACATGCTGAAGCTATCCATACCAAGATGAAATAT 4168  
DB 773 GATGACTTTTAAATGATTAATTAATGATGATTAATGAGAGAGTAAAGAAAT 832  
QY 4169 GATGTTGATGAAATGAGAGTTTGTATATATCATGATGACGACGACGACGACGAC 4228  
DB 833 ATTTGAGAGAGAGCAAGTAAATGACGATTTTAAATGATTAATGATTAATGATTAATGATTAAT 892

QY 4229 TACACATTGAGAAAGAGTTTACGAAAAATTTGCTGCTACTTAACATTCACTTGA 4284  
 DB 893 CAAGAACACACACACATGTTGAAGAAAAAGTTGAAGAAAGTTGAAGAAATCA 948

RESULT 9

US-09-640-173-53/c  
 ; Sequence 53, Application US/09640173  
 ; Patent No. 6613515  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Xu, Jiangchun  
 ; TITLE OF INVENTION: OVARIAN TUMOR SEQUENCES AND  
 ; FILE REFERENCE: 210121.484C2  
 ; CURRENT APPLICATION NUMBER: US/09/640.173  
 ; CURRENT FILING DATE: 2000-08-15  
 ; NUMBER OF SEQ ID NOS: 196  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 53  
 ; LENGTH: 396  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapien  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (1)...(396)  
 ; OTHER INFORMATION: n = A,T,C or G  
 US-09-640-173-53

Query Match 0.9%; Score 55.6; DB 4; Length 396;  
 Best Local Similarity 47.4%; Pred. No. 0.0014;

Matches 118; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

QY 1130 GAAGAAATGTGATCATCTGATATGCTAGATTGTTGACATTCAGACAAATAT 1189  
 DB 304 GAAATATNN 245  
 QY 1190 GAACCTATCTTATGTTAGTTTACACACACAGAGAAACCAAGAAACCGAGAG 1249  
 DB 244 AAGGAAAAAANNN 185  
 QY 1250 AACTCAGTGAACATCTCAAGAGAAAGAACTGAAATCAAGAAAGAAATAGAAAT 1309  
 DB 184 AA 125  
 QY 1310 CCAAAATCAGTACGATTAAGTAAAGGAGAAAAAGAAAAAGAAAAAATTCCTG 1369  
 DB 124 AA 65  
 QY 1370 TGAATAACA 1378  
 DB 64 AAAAAAAAAA 56

RESULT 10  
 US-09-713-550-53/c  
 ; Sequence 53, Application US/09713550  
 ; Patent No. 6617109  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Xu, Jiangchun  
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE  
 ; FILE REFERENCE: 210121.484C4  
 ; CURRENT APPLICATION NUMBER: US/09/713.550  
 ; CURRENT FILING DATE: 2000-11-14  
 ; NUMBER OF SEQ ID NOS: 205  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 53  
 ; LENGTH: 396  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapien

FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (1)...(396)  
 ; OTHER INFORMATION: n = A,T,C or G  
 US-09-713-550-53

Query Match 0.9%; Score 55.6; DB 4; Length 396;  
 Best Local Similarity 47.4%; Pred. No. 0.0014;

Matches 118; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

QY 1130 GAAGAAATGTGATCATCTGATATGCTAGATTGTTGACATTCAGACAAATAT 1189  
 DB 304 GAAATATNN 245  
 QY 1190 GAACCTATCTTATGTTAGTTTACACACACAGAGAAACCAAGAAACCGAGAG 1249  
 DB 244 AAGGAAAAAANNN 185  
 QY 1250 AACTCAGTGAACATCTCAAGAGAAAGAACTGAAATCAAGAAAGAAATAGAAAT 1309  
 DB 184 AA 125  
 QY 1310 CCAAAATCAGTACGATTAAGTAAAGGAGAAAAAGAAAAAGAAAAAATTCCTG 1369  
 DB 124 AA 65  
 QY 1370 TGAATAACA 1378  
 DB 64 AAAAAAAAAA 56

RESULT 11  
 US-09-410-464-9  
 ; Sequence 9, Application US/09410464  
 ; Patent No. 6395892  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Straus et al.  
 ; TITLE OF INVENTION: Floral homeotic genes for manipulation of flowering in  
 ; FILE REFERENCE: 53375  
 ; CURRENT APPLICATION NUMBER: US/09/410.464  
 ; CURRENT FILING DATE: 1999-10-01  
 ; EARLIER APPLICATION NUMBER: 09/287,700  
 ; EARLIER FILING DATE: 1999-04-06  
 ; EARLIER APPLICATION NUMBER: 60/080,851  
 ; NUMBER OF SEQ ID NOS: 24  
 ; SOFTWARE: Patentin Ver. 2.0  
 ; SEQ ID NO 9  
 ; LENGTH: 11485  
 ; TYPE: DNA  
 ; ORGANISM: Populus balsamifera subsp. trichocarpa  
 US-09-410-464-9

Query Match 0.9%; Score 55.6; DB 4; Length 11485;  
 Best Local Similarity 44.6%; Pred. No. 0.0049;

Matches 348; Conservative 0; Mismatches 424; Indels 8; Gaps 3;

QY 4468 CAACAAATCGACTATATGATCCGGAGATGATAAGTCATGATGCTGAACTAGAGA 4527  
 DB 130 CTAATAAGAGGCAATCTTGATCCGCTCGGCAACAACCTAATGAATGAATTTCTGC 189  
 QY 4528 ATTAGATCAAAAGATGTTTACGAGAAAGTTCCAAATTCACCGGTGGAAGCCATATAC 4587  
 DB 190 TTGCAATAGACATCTGAGATCTGGATCTGCTCTACTCCGAGTAAGAGTGTGG 249  
 QY 4588 TATGGGTGGGTACATCTAGAGAAATGATTTCTCTCAAGAGTGTGGTGGAAATACG 4647  
 DB 250 TTGCAATGGGTGTATGATCAAGATCAATATCT--GATGGGTATATGAGCAATACAA 306  
 QY 4648 TTGTGTGTCATGAGCAACAGCAAGAAATGATGATTAAGCCCTTTAGGTGAG 4707  
 DB 307 AGCTAGCGCTTGCAGAAAGATCTCTCAATTAATGATGAGTATGAGAAACATT 366

QY 4708 TTCACTGTATAGATCTTGTGACATTAATGACAAATAATAGTGTGATTAAG 4767  
DB 367 TGCCTCGGTGCAAAAATGACTATCTTCTTATCTTATGCTGAGCTTGCATTA 426  
QY 4768 AATGACATTCACATTTAGACGTGAGTGGCGTATCTAATGCCCTATTACTGATTC 4827  
DB 427 GTGCAATTTCTCAGCTGTATGTTAAAAATGCTTCTTGATGAGATCTTCAGAGAA 486  
QY 4828 AATTCATTTATGTCTTCTCTCTTAATCAGTACCTTTGAGAAAAACATTTGTTGTT 4887  
DB 487 AGTTATGTGCACTCCCTCTGATTTATCATATGACTC--TGGATATGTTTGTAACTT 544  
QY 4888 ATTGAACGTTCTGATGAGGTTTAAACAGTCGCGGTTTGGATGGTATCACAATCA 4947  
DB 545 AAGAAAGCATTAATTAATGTCTCAACAAGACCCCGCTGTGTTGAGAAATTC 604  
QY 4948 AAGAGTATGGAACATTTGTTTACTCAAGTTTACACATGATGTTTATTACAT 5007  
DB 605 TATGTATCTCGCTCTGTCATTTGTTCTAGACAGTATGCTGCTCTTTT--AT 661  
QY 5008 TGAATATGAAGGATCAGTAATATATTAGTTTATGTTGATGATCTCTATGTT 5067  
DB 662 TAAATGACATGATCAGGTGATCATCTCTTATATGTTGATTAACATGATTTAT 721  
QY 5068 TGAAGTTCACAAAAGTATATGATAATTTTGTGATCAATGAGATCATTTTGAAGT 5127  
DB 722 TGGATGATGACATGATGATGATTTTCACTTGAAGACAAAGTGGCTAGACATTTGA 781  
QY 5128 TAAAGTGTGTTGTAATATCAATATCTTGTATGTAATTTGTAACCAATCTGG 5187  
DB 782 GAAAGATTTGGTTATCTTCAATATTTCTGGGTAATGAGTACATCTCACCCTAGAG 841  
QY 5188 TTATATTTTATCTCAAGAAAAATTTCTCAAGAAATTAATTAGATTTCAACTAGTGA 5247  
DB 842 TTACTCTCTCTCTAGTGAATAATGTTGAGATATTCTGAGACAGCTAGACTTACTGA 901

RESULT 12  
US-09-254-776B-50  
; Sequence 50, Application US/09254776B  
; Patent No. 6559359  
; GENERAL INFORMATION:  
; APPLICANT: latech, Howard  
; TITLE OF INVENTION: PLANT RETROVIRAL POLYNUCLEOTIDES AND METHODS FOR USE THEREOF  
; FILE REFERENCE: 27013/33479A  
; CURRENT APPLICATION NUMBER: US/09/254, 776B  
; CURRENT FILING DATE: 1999-03-09  
; NUMBER OF SEQ ID NOS: 86  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 50  
; LENGTH: 2826  
; TYPE: DNA  
; ORGANISM: Glycine max  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: SIRE-1 genomic clone  
US-09-254-776B-50

Query Match 0.9%; Score 55.4; DB 4; Length 2826;  
Best Local Similarity 43.7%; Pred. No. 0.0032;  
Matches 343; Conservative 0; Mismatches 436; Indels 6; Gaps 2;

QY 4484 AATGATCCGGGATGATTAAGTCATGATGCTGAACATGAGAAATTTAGATCAAAAGAT 4543  
DB 1324 ACTGATGATTTCTGATCAATGCTATGCAAGAAATGAGCAATTCAAAAGAAATGAA 1383  
QY 4544 GTTTAAGAGAAAGTTCATTTCCACCGGTGGAAGCTATATCTATGAGTGTGGTACAT 4603  
DB 1384 GTTTGGAGCTAGTTCCTAGGCCCGAGGAACTATATGATGAGCACCACAGTGCATCTTC 1443  
QY 4604 ACTGAGAAATGATTTCTCTCAAGTGTGTTGGAAATACAGTGTGTTGTCATGAGGC 4663

DB 1444 AAGAAACAAACAAAT--GAAAGAGTGTATATACCAAGAAACAAGCCAGACTTGTGCT 1500  
QY 4664 AACAGACAAAAGAAAAATTTGATTTATGACCTTTTATGTTAGTTACCTGTTATATGAT 4723  
DB 1501 CAAGGCTACATCAGATTTGAAGGTGATGACTTGTATGATAAACTTTGCCCTGGTGTAA 1560  
QY 4724 CTGTGACTATTAATTTATGACATTAATGTTGTATTAATTTAGGAATGACAAATTCACAT 4783  
DB 1561 CTGATGTCATCAACACTGTTACTGTTGTGAGCTTGATCTCAATTTCAAGCTGTACAG 1620  
QY 4784 TTAGACGTGAGTGGCGGTATCTAATATGCCCTATTAATCTATCAATTCAAATTTATGTC 4843  
DB 1621 ATGATGTGAAGACGACATTTCTGATGATGATGATGATGATGATGATGATGATGATG 1680  
QY 4844 TTTCTCTTAATATGATGATCTTTGAGAAAAACATTTGTTTATGAAACGTTCTGTC 4903  
DB 1681 CCAAGGATTTGATGATCTCAACTATCCAGATATGATATGATGATGATGATGATGATGATG 1740  
QY 4904 TATGGTTTAAACAGTGGGTTTGAATGATGATGATGATGATGATGATGATGATGATGATG 4963  
DB 1741 TATGATTTGAAGCAAGCTTCAAGAGCTTGTATGATGATGATGATGATGATGATGATGATG 1800  
QY 4964 ATGAGTTTACTCAAGTTTACACATGATGATGATGATGATGATGATGATGATGATGATG 5023  
DB 1801 CAAGGATTTAGAAAGGGGGGATGATGATGATGATGATGATGATGATGATGATGATGATG 1857  
QY 5024 TCAGTAATATTTATGATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5083  
DB 1858 AATG 1917  
QY 5084 GTTATGATTAATTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5143  
DB 1918 ATGCTTCAACATTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1977  
QY 5144 ATATCAATTTATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5203  
DB 1978 CTGAATTTATTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2037  
QY 5204 GAAATATTTCTCAAGAAATTTACTTAAGGATTTCAACTGATGATGATGATGATGATGATG 5263  
DB 2038 AGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2097  
QY 5264 ATACC 5268  
DB 2098 ACACC 2102

RESULT 13  
US-09-835-811-1  
; Sequence 1, Application US/09835811  
; Patent No. 6482936  
; GENERAL INFORMATION:  
; APPLICANT: HU, Song et al  
; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,  
; FILE REFERENCE: CL0012228  
; CURRENT APPLICATION NUMBER: US/09/835, 811  
; CURRENT FILING DATE: 2001-04-17  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 1696  
; TYPE: DNA  
; ORGANISM: Human  
US-09-835-811-1

Query Match 0.9%; Score 55.2; DB 4; Length 1696;  
Best Local Similarity 57.6%; Pred. No. 0.0029;  
Matches 99; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 1207 TATGTTACCAAGCAAGAGAAACCAAAAGAGAAACCAAGAGAACTCACTGGAACATTC 1266

Db 1506 TAGATTAAA 1567  
QY 1267 TCAAAAGAGAACTGAAATCAGAAAGAAATPAGAAACATCCAAATCAGTAAACA 1326  
Db 1568 AA 1627  
QY 1367 TAAAGTGAAAAAGAAAAAGAAAAAACTTCACTGAGATGAAAAACA 1378  
Db 1628 AA 1679

RESULT 14  
US-08-771-602D-1  
Sequence 1, Application US/08771602D  
Patent No. 5926795  
GENERAL INFORMATION:  
APPLICANT: Voytas, Daniel F.  
APPLICANT: Zou, Sige  
TITLE OF INVENTION: Retrotransposon and Methods  
NUMBER OF SEQUENCES: 51  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Greenlee, Winner and Sullivan, P.C.  
STREET: 5370 Manhattan Circle, Suite 201  
CITY: Boulder  
STATE: Colorado  
COUNTRY: USA  
ZIP: 80303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/771.602D  
FILING DATE: 20-DEC-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/010,869  
FILING DATE: 31-JAN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Feiber, Donna M.  
REGISTRATION NUMBER: 33,878  
REFERENCE/DOCKET NUMBER: 8-96  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 499-8080  
TELEFAX: (303) 499-8089  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6660 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Saccharomyces paradoxus  
STRAIN: NRRL Y-1717  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1441..6321  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 1228..6602  
OTHER INFORMATION: /function= "retrotransposon"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 1228..1478  
OTHER INFORMATION: /function= "5' LTR of Ty5-6p"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 6352..6602

OTHER INFORMATION: /function= "3' LTR of Ty5-6p"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 2852..4827  
OTHER INFORMATION: /function= "integrase region coding  
OTHER INFORMATION: region of Ty5-6p"  
US-08-771-602D-1

Query Match 0.8%; Score 54.6; DB 2; Length 6660;  
Best Local Similarity 45.2%; Pred. No. 0.0067;  
Matches 241; Conservative 0; Mismatches 289; Indels 3; Gaps 1;

QY 4661 GGCAACAGCAAAAGAAAAAATGATATGACCCCTTTAGTGTAGTTACCTGTATA 4720  
Db 4981 GCCCAAGACATACCTAAAGAGCTGTATGACTATCAAGAACTTTGACCACTGAT 5040  
QY 4721 GATCTTGACTATTAAGATTATGACAAATATAGTTGAAATTGAAATGACAAATCAA 4780  
Db 5041 CGATATGACTCTGTATGATTATTTCTGCGCTGTAGCTGCTCAAACTAATAGTATAT 5100  
QY 4781 CATTTAGCTGAGTGGCGGCTATCTAATGCTCTATATCTCAATTCATTTAT 4840  
Db 5101 CGATGACGTTGACACCGCGTTTCTAATCTAATAATGATGACCGGTATACGTAA 5160  
QY 4841 GTCTTCTCTCTAAATCAGTACCTTTGAAAGAAAAACATTTGTTATTTGAAAGCTTCT 4900  
Db 5161 CAACACCGGATTTATTAATGAAGTATCCCGACTATGTAGGAATATACGCGGT 5220  
QY 4901 GTCTATGGTTAAACAGTCGGGTTTGAATGTATACACATATCAAAAGATTTGCA 4960  
Db 5221 ATGTATGACTCAGAGAACCCCATTAATGAAACGAAATATCAATATCTCTTCA 5280  
QY 4961 GACATGTTTACTCAATTTTACAAATGATGATTTTACATTTGATGAAATGAAAG 5020  
Db 5281 AGATTGTTTTCGTGATGATGAGCGGACATGGCTTAT--ATTTCGTTCCATCT 5337  
QY 5021 GGATCAGTAATATTTAGGTTTATATGTTGATGATTTCTTATGTTGAAAGTTCA 5080  
Db 5338 GATGTCCTCATCTACCTGCCCCATATAGTACGATGATCTTGTGCTGCCCTCCG 5397  
QY 5081 AAAGTATGATATTTTGTGATCAATGAGAGATCAATTTGAAGTTAAAGTTTGT 5140  
Db 5398 AATATATGACAGGTTTATGACAAACTAAGAACTTATCTCATGAGATCTAGGT 5457  
QY 5141 GAAATATCAATTTATCTTGTTATGATTTCTGTAACGAAATCTGTTATAT 5193  
Db 5458 AAAGTTGACAAATTCCTCGTCTTAACATTAATCAATTTTCAATGAGAGAT 5510

RESULT 15  
US-09-232-446B-1  
Sequence 1, Application US/09232446B  
Patent No. 6228647  
GENERAL INFORMATION:  
APPLICANT: Voytas, Daniel F.  
APPLICANT: Gal, Xisewu  
TITLE OF INVENTION: Transposable Element Protein that Directs DNA  
INTEGRATION TO SPECIFIC CHROMOSOMAL SITES  
FILE REFERENCE: 2-98  
CURRENT APPLICATION NUMBER: US/09/232,446B  
CURRENT FILING DATE: 1999-01-15  
PRIOR APPLICATION NUMBER: US 60/071,383  
PRIOR FILING DATE: 1998-01-15  
NUMBER OF SEQ ID NOS: 26  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 1  
TYPE: DNA  
LENGTH: 6660  
ORGANISM: Saccharomyces paradoxus  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1441)..(6318)  
US-09-232-446B-1

Query Match 0.8%; Score 54.6; DB 3; Length 6660;  
Best Local Similarity 45.2%; Pred. No. 0.0067;  
Matches 241; Conservative 0; Mismatches 289; Indels 3; Gaps 1;

Qy	4661 GGCACGACAAAGGAAATGGATTATGACCCCTTTAGTGTAGTTCACCTGTATA	4720
Db	4981 GCCCAAGACATACCTCAAAAGCTGATGACTATCAGAACTTTGACCCAGTCATT	5040
Qy	4721 GATCTGTGACTATATAGATTATGACAAATATAGTGTGCAATAGCAATTCAA	4780
Db	5041 CGATATGACTGTGTAATATTTCTGCCCCTGTAGCTGCCCTCAACTAATAGTATAT	5100
Qy	4781 CATTAGACGTGAGTGGCGTATCTAAATGCTCTATTACTATTCAATTCATTTAT	4840
Db	5101 CAGATGACCGTTGACACCGGTTCTAACTCAAAAATGATAGCCGGTATAGTAA	5160
Qy	4841 GTCCTTCCTCTAATACAGTACCTTGAGAGAAAACCATGTGTTATGAAACGTCT	4900
Db	5161 CAACCAACCGGATTTATTAAGAAATATCCGACTATGTATGGAACTATACGGCGGT	5220
Qy	4901 GTCATGGGTTAAACAAGTGGGTTTGGAAATGTTATCACTATCAAAAGATTTGAA	4960
Db	5221 ATGTATGGACTCAAGCAAGCCCCCATCTACTATGGAACGACATATCAACATCTCTTCAA	5280
Qy	4961 GACATTTGTTTACTCAAGTTTACACAATGATGTTTATTCACATTTGAATGAAG	5020
Db	5281 AAGATGGTTTTCGTGACATGAAGCGCAACATGGCTTAT--ACTTTCGTTCCACATCT	5337
Qy	5021 GGATCAGTATATATTTAGTTTATATGTATGATGATATTTCTTATGTTGAAGTTCA	5080
Db	5337 GATGGTCCCATCTACATTTGCCCTATACGTAGACGACTTACTGTGTCTCCCTCCCG	5397
Qy	5081 AAAGTTATTTGATTTTGTGATCATTTGAGATCATTTTGAAGTTAAGGTTTGGT	5140
Db	5398 AAAATATATGACAGGTTAAGCAGAACTAAGAGTTATCTCAATGAAGATCTAGGT	5457
Qy	5141 GAATATCAATTTATCTTGTATGAAATTCGTAACCAATCTGTTATAT	5193
Db	5458 AAAGTTGACAAATTCCTGCTTAACTTATCATTTTCAATGAGACAT	5510

Search completed: April 11, 2004, 21:18:25  
Job time : 294 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 11, 2004, 18:29:18 ; Search time 1445 Seconds

(without alignments)  
16682.970 Million cell updates/sec

Title: US-09-430-590E-3

Perfect score: 6426

Sequence: 1 tgggtgcttgacacattt.....agaagttatctccatca 6426

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 2475585 seqs, 1875730760 residues

Total number of hits satisfying chosen parameters: 4951170

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:\*

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*

2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*

3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*

4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*

5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*

6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*

7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*

8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*

9: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*

10: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*

11: /cgn2\_6/ptodata/2/pubpna/US09C\_NEW\_PUB.seq:\*

12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*

13: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*

14: /cgn2\_6/ptodata/2/pubpna/US10C\_NEW\_PUB.seq:\*

15: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*

16: /cgn2\_6/ptodata/2/pubpna/US10C\_NEW\_PUB.seq:\*

17: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

18: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	92	1.4	1289	US-10-424-599-71662	Sequence 71662, A
2	89.8	1.4	5417	US-10-424-599-103057	Sequence 103057, A
3	84.2	1.3	4029	US-10-424-599-42489	Sequence 42489, A
4	84.2	1.3	4143	US-10-424-599-42491	Sequence 42491, A
5	83.8	1.3	13341	US-08-910-386A-1	Sequence 1, Appl1
6	79.6	1.2	78333	US-10-298-122-3	Sequence 3, Appl1
7	70.8	1.1	560	US-10-424-599-118696	Sequence 118696, A
8	70.6	1.1	2760	US-10-424-599-108338	Sequence 108338, A
9	68.4	1.1	3673778	US-10-312-841-1	Sequence 1, Appl1
10	68.2	1.1	7305	US-10-311-455-1610	Sequence 1610, Appl1
11	65.8	1.0	3673778	US-10-312-841-2	Sequence 2, Appl1
12	65.6	1.0	7628	US-10-424-599-108640	Sequence 108640, A
13	65.4	1.0	11729	US-10-311-455-868	Sequence 868, Appl
14	65.2	1.0	681	US-09-822-830A-26	Sequence 26, Appl
15	65.2	1.0	12177	US-10-311-455-624	Sequence 624, Appl

16	65	1.0	628	US-10-424-599-62262	Sequence 62262, A
17	65	1.0	2685	US-10-424-599-108648	Sequence 108648, A
18	64.8	1.0	14798	US-10-311-455-1005	Sequence 1005, Ap
19	63.4	1.0	65	US-10-032-585-2429	Sequence 2429, Ap
20	63.4	1.0	9358	US-10-396-122-90	Sequence 90, Appl
21	63.4	1.0	24259	US-10-221-714A-415	Sequence 415, Ap
22	63.2	1.0	2180	US-10-424-599-9803	Sequence 9803, Ap
23	63.2	1.0	3861	US-10-424-599-8892	Sequence 8892, Ap
24	63.2	1.0	3879	US-10-424-599-8903	Sequence 8903, Ap
25	63.2	1.0	3912	US-10-424-599-8914	Sequence 8914, Ap
26	63.2	1.0	335913	US-09-754-853A-2	Sequence 2, Appl1
27	63.2	1.0	335913	US-09-754-853A-3	Sequence 3, Appl1
28	62.8	1.0	6192	US-10-221-714A-408	Sequence 408, Ap
29	62.4	1.0	8946	US-10-311-455-884	Sequence 884, Ap
30	62	1.0	767	US-10-424-599-45104	Sequence 45104, A
31	62	1.0	9515	US-10-239-676-160	Sequence 160, Ap
32	62	1.0	9515	US-10-240-453-182	Sequence 182, Ap
33	62	1.0	10369	US-10-221-714A-25	Sequence 25, Appl
34	62	1.0	10369	US-10-311-455-365	Sequence 365, Ap
35	61.8	1.0	1460	US-10-424-599-142655	Sequence 142655, A
36	61.4	1.0	327	US-09-960-352-4630	Sequence 4630, Ap
37	61.4	1.0	799	US-10-424-599-88818	Sequence 88818, A
38	61.4	1.0	4985	US-10-094-240-10	Sequence 10, Appl
39	61.4	1.0	4985	US-10-056-405-10	Sequence 10, Appl
40	61.2	1.0	1198	US-09-789-561-55	Sequence 55, Appl
41	61	0.9	1113	US-10-083-357-599	Sequence 599, Ap
42	61	0.9	2973	US-10-425-114-1139	Sequence 1139, A
43	61	0.9	2961	US-10-424-599-58746	Sequence 58746, A
44	61	0.9	9399	US-10-396-122-93	Sequence 93, Appl
45	60.8	0.9	973	US-10-424-599-58879	Sequence 58879, A

#### ALIGNMENTS

RESULT 1

US-10-424-599-71662

Sequence 71662, Application US//10424599

Publication No. US20040031072A1

GENERAL INFORMATION:

APPLICANT: La Rosa Thomas J

APPLICANT: Kovalic David K

APPLICANT: Zhou Yinhua

APPLICANT: Cao Yongwei

TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with

TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

FILE REFERENCE: 38-21(53223)B

CURRENT APPLICATION NUMBER: US/10/424,599

CURRENT FILING DATE: 2003-04-28

NUMBER OF SEQ ID NOS: 285684

SEQ ID NO 71662

LENGTH: 1289

TYPE: DNA

ORGANISM: Glycine max

FEATURE:

NAME/KEY: unsure

LOCATION: (1)..(1289)

OTHER INFORMATION: unsure at all n locations

FEATURE:

OTHER INFORMATION: Clone ID: PAT\_MRT3847\_35724C.1

US-10-424-599-71662

Query Match

Best Local Similarity 46.3%; Pred. No. 5.5e-08;

Matches 376; Conservative 0; Mismatches 430; Indels 6; Gaps 2;

QY 4486 TGAATCCGGGATGATTAAGCAATGCAATGCTGACACAGAAATTAGTCAAAAGATGT 4545

Db 297 TGAATCAAAATGAGAAATGCAATGAAAGAGAGCTTAATGATTTGAAAAAATAAATTAAC 356

QY 4546 TTAAGAGAGATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTC 4605

Db 357 ATGGAGAGCTGTGTGACAGACCTTAACCAAGAAAGAGATTTGTGTGATGATGATTTATATAG 416

```

QY 4606 TGAGAAATGATCTCTCAAAAGGTGTTGGAAATACGTTGTTGCCATGSCAA 4665
DB 417 AACCAAGCTCAATTC---AGATGCTCTATAAACAATACAGCAAGGCTTGTGTTAA 473
QY 4666 CAGCAAAAAGAAAAATTTGATTTGACCTTTTACCTTTAGTTCACCTGTTATATCT 4725
DB 474 AGGATATGCTCAAGTGTGTTGGTGTGTTTTCAGAGACTTTTGCCTTACCTGCTCACT 533
QY 4726 TGTACTATATAGATTTATTTGACAAATATATAGTTGATTTAGATGACATTCACATTT 4785
DB 534 AGAAACCTTAAACAGAGGCGCGCTCTTGTGACCAAAAAGGTGATTTATACATCAAAAT 593
QY 4786 AGAGCTGAGTCGGCGGTATCTAAATGCTCTATTTACTTCAATTCATTTATGCTT 4845
DB 594 GGATGTGAATACACCTCTTAAATGATCTTGAGAGAAATTTTGTGACGACACC 653
QY 4846 TCTCTTAATCACTACCTTTGACAAATATGTTGTTATTTGAAACGTTCTGCTA 4905
DB 654 TGAAGGTTGCTGTTGTTACAGACAGAGAGAAATATGATTTAGATTAAGAAAGCTTTGTA 713
QY 4906 TGGGTTAAACAGTCGGGTTTGAATGATTCACACTATCAAAAAGATTTGAAAGCAT 4965
DB 714 TGGTTAAAGCAGCGCACCGCGCTATGATGATGATGATGATGATGATGATGATGATGAT 773
QY 4966 TGGTTTACTCAAGTTTATACCAATGATGATGATGATGATGATGATGATGATGATGAT 5025
DB 774 AGGCTTCAAAAAGTTTATGATGATCTATTTATGATGATGATGATGATGATGATGATGAT 833
QY 5026 AGTAAATATTTAGGTTTATATGTTGATGATATTTCTTATGTTGAAAGTTTCAAAAAAGT 5085
DB 834 GCTTGTGTA---GTATCTCATATGTTGATGATCTATGTTGCGGAAAGCAGCGGAGCA 890
QY 5086 TATGATATTTTGTGATCAATTGAGAGATCACTTTGAGTTTAAAGTTTGTGTAAT 5145
DB 891 TATGACAAAGTTCAAGAGAAAGAAAGATGCTTTGAAATGACACCTTTGAAAGAT 950
QY 5146 ATCAATATCTTGTGATTTGATTTGTTAAACCGAATCGGTTATTTTATCTCAGA 5205
DB 951 GACATCTTCTTCGATGATGAGTGTGATCAAAAACAGAAATATTTCTTGGCAGGA 1010
QY 5206 AAAATTTCTCAAAATTTACTTTAAGATTTCAAACTGATGATCAATGAGAAAAAT 5265
DB 1011 AAAATATGCAATGAAATTTCTCAAGAGATTTCAACATGAAATGTTGCAAAACCACTGCAC 1070
QY 5266 ACCCGATTCGATGACAAATATGAAAAGG 5297
DB 1071 TCCATGATCAAAAAGAGAGTTTACAAAG 1102

```

```

RESULT 2
US-10-424-599-103057
; Sequence 103057, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongsuei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 103057
; LENGTH: 5417
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_64078C.1
US-10-424-599-103057

```

```

Query Match 1.4%; Score 89.8; DB 12; Length 5417;
Best Local Similarity 46.8%; Pred. No. 3e-07; Indels 0; Gaps 0;
Matches 283; Conservative 0; Mismatches 322;
QY 4664 AACAGACAAAAGAAAAATTTGATTTAGACCTTTTATGTTTACTTACCTGTTATAGAT 4723
DB 3760 AAAGTTTAAAGCAACATATGAGTGTATATGATGAAAGTTTGCACCGGTTGCCCGC 3819
QY 4724 CTTCGATTTAAGATTTTTCACATTAATAGTGTGATTTAGAAATGACATTCACAT 4783
DB 3820 ATGAAACCATTCGCTCTTATTTCTTGGACCTCAAAAGAGTGAATTTTTCAG 3879
QY 4784 TTAGAGTCAGTCGGCGGTATCTAAATGCTCTATTTACTATTCATTCATTCATTTATGTC 4843
DB 3880 CTGATGTTAAATGCGCATTTCTAAATGCGTATTTGAAAGAAATGCTATGTTGAACAA 3939
QY 4844 TTTCCTCTTAATCAGTACCTTTGAGAAAAACATTTGTTTATTTGAAACGTTCTGTC 4903
DB 3940 CCAATGGGTTTGTATGAAAGTCAAGAAAGAAAGTCTGAAATTTGAACAGGCGTTG 3999
QY 4904 TATGGTTTAAACAGTCGGGTTTGAATGATCACTATCAAAAAGATTTGAAAGAC 4963
DB 4000 TATGTTTAAAGACAGCACCGAGGATGAAATCTGCAATGACAGTATCTTCCAGAC 4059
QY 4964 ATGTTTACTCAAGTTTTCACAAATGATGTTTATTTTCACTGATTAATGAAAGAGGA 5023
DB 4060 AATGGTTTGTTCGTTGCAAAATGAGTATCTCTTATGTTAAACCTTTTAAATATGAT 4119
QY 5024 TCAGTAAATATTTAGGTTTATATGTTGATGATATTTCTTATGTTGAGTTCACAAA 5083
DB 4120 GATGCTTATTTATTTGCTTTATGTTGATGATGATGATGATGATGATGATGATGATGAT 4179
QY 5084 GTTATGATTAATTTTGTGATCAATTTGAGATCACTTTTGAAGTTTAAAGTTTGTGAA 5143
DB 4180 TTGTTTAAAGACTTCAAGAGTCCATGCTGTAATTTGATGACAGATATGAGACTC 4239
QY 5144 ATATCAATTTCTTGTGATTTGAAATTTGTTAAACCGAATCGGTTATTTTATCTCA 5203
DB 4240 ATGCTATTTACTTGGGATGAAAGTGAAGCAAAACGGAATGATCTTGTCTCACAA 4299
QY 5204 GAAATTTCTCAAGAAATTTACTTATGATTTCAAACTGATGATGATGATGATGATGATGAT 5263
DB 4300 GAAAGTACAAAGAAAGTGTGAAAGAAATTTAATATGCTTATGTAATCCCGTGAAC 4359
QY 5264 ATACC 5268
DB 4360 ACACC 4364

```

```

RESULT 3
US-10-424-599-42489
; Sequence 42489, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongsuei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 42489
; LENGTH: 4029
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_138369C.1
US-10-424-599-42489
Query Match 1.3%; Score 84.2; DB 12; Length 4029;

```

Best Local Similarity 46.3%; Pred. No. 3.3e-06;  
Matches 352; Conservative 0; Mismatches 403; Indels 6; Gaps 2;

```
QY 4496 TGGATAAGTCAATGATGCTGATGATGAAATTTAATGAAAGATGTTAGAGAA 4555
DB |||||
QY 2521 TGGCTGAAGCTATGAAAGAGAACTGATCCATGAGAAAGAACTTGAGAAATTG 2580
DB |||||
QY 4556 GTTCCAAATTCACCGGTGTGAAGCTATATCTATGAGTTGGGTGATCACTAGAGAAATT 4615
DB |||||
QY 2581 GTTGATTTACCAAAAGAAAGAAAGCAATAGGTGTGAATGGGTGTAAAGTAAAGTT 2640
DB |||||
QY 4616 GATTCTCAGAAAGGTGTGTGGGAAATCAGTGTGTGTCATGCGAAGCAAG 4675
DB |||||
QY 2641 AATCCT--AAAGTGAAGTGTGAAGTGAAGCAAGCAAGTGTAGTGAAGATTCCTC 2697
DB |||||
QY 4676 GAAAAATTGATGATGACCTTTTGTAGTTCACCTGTATAGATCTTGACTATA 4735
DB |||||
QY 2698 CAAGACATGCTGTGATGATGAGAAATTTTGACCTGTGCAAGATTTGAAGTGTG 2757
DB |||||
QY 4736 AGATTATTGACATATATAGTTGTGATGAGAAATGACAAATTCACATTTAGCTGAG 4795
DB |||||
QY 2758 AGAGTTGTGTAGCTCATGCTAGTATGCAAGGTGAAGTACATCACTGATGATAA 2817
DB |||||
QY 4796 TCGGCTATCTTAATGAGAAATTTGAAAGCAATCCAAATTTATGCTTCTCCTAAA 4855
DB |||||
QY 2818 TCTGCTCTTATATGAGAAATTTGAAAGCAATTTATGAGCAAGCAAGTGTG 2877
DB |||||
QY 4856 TCAGTACCTTTGAGAAAGAAACCATTTGTGTATGAAACGTTCTGTCTATGGTTAAA 4915
DB |||||
QY 2878 GTGAACAGAGAAAGTGAAGAAAGTGTGAAATTTGAAAGCAAGCTTGTATGCTGAAA 2937
DB |||||
QY 4916 CAGTGGGTTTGGATGATGATGACATCTATCAAAAGATTTGAAAGCAATTTGTTTACT 4975
DB |||||
QY 2938 CAGGTCACAGCTTGAAACAAAGATGATGAAAGTGTATGCAATGAGGTGTATGTG 2997
DB |||||
QY 4976 CAAGTTTACACATGATGTTTATTTTCACTTGAATGAAAGGATCTGATATAT 5035
DB |||||
QY 2998 AGATGATATTTGAGACATGAGAGTGTATGTAAGTCAAGGGGTGTGATATGTAATA-- 3054
DB |||||
QY 5036 TTAGGTTATATGTTGATGATATTTCTTATGTTGGAATTCACAAAAGTTATGATAT 5095
DB |||||
QY 3056 CTATGCTCTATGTTGATGATGTTGTGTCTGTAACAAATGAACTGATGATCTGAA 3114
DB |||||
QY 5096 TTGTGATGATGATGAGATGATTTGAGTTAAAGTGTGAGAAATTCAAATAT 5155
DB |||||
QY 3115 TTTAAAGAAATATGATGATGATCTTTGAAATGACAGCTGAGAGATATCTACTTT 3174
DB |||||
QY 5156 CTGATGATGATGATTTGTAACCCGATCTGTTATATTTATCTCAAGAAATTTCTC 5215
DB |||||
QY 3175 CTAGGATGATGATTTCAAAATCTAGTGTGTGCTACTGTCAAGAAAGATATGTT 3234
DB |||||
QY 5216 AAGAAATTTACTTAAGGATTTCAAACTAGATGATGATGATG 5256
DB |||||
QY 3235 TCTGATGATGATGAGATTCATATGCAACCTGTAATG 3275
DB |||||
```

## RESULT 4

US-10-424-599-42491  
; Sequence 42491, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 42491  
; LENGTH: 4143

TYPE: DNA  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRN3847\_138370C.1  
US-10-424-599-42491

Query Match 1.3%; Score 84.2; DB 12; Length 4143;  
Best Local Similarity 46.3%; Pred. No. 3.4e-06;  
Matches 352; Conservative 0; Mismatches 403; Indels 6; Gaps 2;

```
QY 4496 TGGATAAGTCAATGATGCTGATGATGAAATTTAATGAAAGATGTTAGAGAA 4555
DB |||||
QY 2635 TGGCTGAAGCTATGAAAGAGAACTGATCCATGAGAAAGAACTTGAGAAATTG 2694
DB |||||
QY 4556 GTTCCAAATTCACCGGTGTGAAGCTATATCTATGAGTTGGGTGATCACTAGAGAAATT 4615
DB |||||
QY 2695 GTTGAAATTTACAAAGAAAGAAAGCAATAGGTGTGAATGGGTGTAAAGTAAAGTT 2754
DB |||||
QY 4616 GATTCTCAGAAAGGTGTGTGGGAAATCAGTGTGTGTCATGCGAAGCAAGCAAG 4675
DB |||||
QY 2755 AATCCT--AAAGTGAAGTGTGAAGTGAAGCAAGCAAGTGTAGTGAAGATTCCTC 2811
DB |||||
QY 4676 GAAAAATTGATGATGACCTTTTGTAGTTCACCTGTATAGATCTTGACTATA 4735
DB |||||
QY 2812 CAAGACATGCTGTGATGATGAGAAATTTTGACCTGTGCAAGATTTGAAGTGTG 2871
DB |||||
QY 4736 AGATTATTGACATATATAGTTGTGATGAGAAATGACAAATTCACATTTAGACGTGAG 4795
DB |||||
QY 2872 AGAGTTGTGTAGCTCATGCTAGTATGCAAGGTGAAGTACATCACTGATGATAA 2931
DB |||||
QY 4796 TCGGCTATCTTAATGAGAAATTTGAAAGCAATCCAAATTTATGCTTCTCCTAAA 4855
DB |||||
QY 2932 TCTGCTCTTATATGAGAAATTTGAAAGCAATTTATGAGCAAGCAAGTGTG 2991
DB |||||
QY 4856 TCAGTACCTTTGAGAAAGAAACCATTTGTGTATGAAACGTTCTGTCTATGGTTAAA 4915
DB |||||
QY 2992 GTGAACAGAGAAAGTGAAGAAAGTGTGAAATTTGAAAGCAAGCTTGTATGCTGAAA 3051
DB |||||
QY 4916 CAGTGGGTTTGGATGATGATGACATCTATCAAAAGATTTGAAAGCAATTTGTTTACT 4975
DB |||||
QY 3052 CAGGTCACAGCTTGAAACAAAGATGATGAAAGTGTATGCAATGAGGTGTATGTG 3111
DB |||||
QY 4976 CAAGTTTACACATGATGTTTATTTTCACTTGAATGAAAGGATCACTGATATAT 5035
DB |||||
QY 3112 AGATGATATTTGAGACATGAGAGTGTATGTAAGTCAAGGGGTGTGATATGTAATA-- 3168
DB |||||
QY 5036 TTAGGTTATATGTTGATGATATTTCTTATGTTGGAATTCACAAAAGTTATGATAT 5095
DB |||||
QY 3169 CTATGCTCTATGTTGATGATGTTGTGTCTGTAACAAATGAACTGATGATCTGAA 3228
DB |||||
QY 5096 TTGTGATGATGATGAGATGATTTGAGTTAAAGTGTGAGAAATTCAAATAT 5155
DB |||||
QY 3229 TTTAAAGAAATATGATGATGATCTTTGAAATGACAGCTGAGAGATATCTACTTT 3288
DB |||||
QY 5156 CTGATGATGATGATTTGTAACCCGATCTGTTATATTTATCTCAAGAAATTTCTC 5215
DB |||||
QY 3289 CTAGGATGATGATTTCAAAATCTAGTGTGTGCTACTGTCAAGAAAGATATGTT 3348
DB |||||
QY 5216 AAGAAATTTACTTAAGGATTTCAAACTAGATGATGATGATG 5256
DB |||||
QY 3349 TCTGATGATGATGAGATTCATATGCAACCTGTAATG 3389
DB |||||
```

## RESULT 5

US-08-910-386A-1  
; Sequence 1, Application US/08910386A  
; Publication No. US20020082041A1  
; GENERAL INFORMATION:  
; APPLICANT: Ronald, Pamela C.  
; APPLICANT: Wang, Guo-Liang  
; APPLICANT: Song, Wen-Yuang  
; APPLICANT: Hulbert, Scot  
; APPLICANT: Richter, Todd

TITLE OF INVENTION: Procedures and Materials for Confering  
NUMBER OF SEQUENCES: 53  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/910,386A  
FILING DATE: 13-AUG-1997  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Basfian, Kevin L.  
REGISTRATION NUMBER: 34,774  
REFERENCE/DOCKET NUMBER: 023070-058950US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 13341 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: Oryza longistaminata  
STRAIN: IRBB21  
POSITION IN GENOME:  
CHROMOSOME/SEGMENT: 11  
MAP POSITION: 11q, RG103  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 2367..4205  
OTHER INFORMATION: /product= "receptor kinase-like protein"  
OTHER INFORMATION: /note= "Xa21 gene family member D"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 4201..5071  
OTHER INFORMATION: /note= "retrofit, a copia-like,  
OTHER INFORMATION: transposon-like element"  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 4484..8821  
OTHER INFORMATION: /product= "retrofit"  
OTHER INFORMATION: /gene= "gag/pol"  
FEATURE:  
NAME/KEY: intron  
LOCATION: 9915..11712  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 10020..10975  
OTHER INFORMATION: /note= "Krispie, transposon-like  
OTHER INFORMATION: element"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 12626..12750  
OTHER INFORMATION: /note= "Pop-O12, transposon-like  
OTHER INFORMATION: element"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 13040..13248  
OTHER INFORMATION: /note= "Ds-rice2, transposon-like  
OTHER INFORMATION: element"  
US-08-910-386A-1

Query Match 1.3%; Score 83.8; DB 8; Length 13341;  
Best Local Similarity 46.4%; Pred. No. 7.1e-06;  
Matches 345; Conservative 0; Mismatches 392; Indels 6; Gaps 2;

QY	4456	TGGATTAAGTCAATGATGCTGAACTGAGAAATTTGATCAAAAGATGTTACGAGAA	4555
DB	7328	TGGAAGATGCAATGCAAACTGATTAATGCTTTGATTAATAATGACATGACACCTA	7387
QY	4556	GTTCCATTTCCACCCGGTGTAGACCTATATCATATGAGTGGGTGACATGCAAAAT	4615
DB	7388	GTTCCATATGAGAAAGCAAAATATCATTTGGGCTGTAATGGATATTAAGATTAAGG	7447
QY	4616	GATTCCTCAAGGCTGTTGTCGGAATTCAGTGTGTGTCATGCGCAACAGCAAAAG	4675
DB	7448	AAGCAGATGGGACACTGATAGATACAAAGCTAGACTGTG---AGCAAAAGGGCTTTAA	7504
QY	4676	GAATAATGATATGACCTTTAGTGTGCTACCTGTATATAGATCTGTGACTATA	4735
DB	7505	CAAAATATGATGATGATTAAGAAATCTTTAGTCTGTTGTTAAAGCTGCTACTATT	7564
QY	4736	AGATTATGACAAATATAGTGTGTAATTAGAAATGACAAATTCACATTTAGACGTGAG	4795
DB	7565	AGATTATGCTGTCATGCTGCTCTAGAGGTGAGTCTTAGACAGTATGATGTGACG	7624
QY	4796	TGGCGGTATCTAATATGCTCTTATCTATCTATCAATCCAAATTTATGCTTCCCTGAAA	4855
DB	7625	AATGCCCTTCTTCATGATGCTTTAGAAAGAGATCTACATGCAACAACTCTCGGCTT	7684
QY	4856	TCAGTACCTTGAAGAAACCATGTTGTTATGAAACGTTCTGCTATGAGGTAA	4915
DB	7686	GAGTATCTCTTAACCTGATTAATGATATGATTAATGATTAAGCATTAATAGGCTGAAA	7744
QY	4916	CAGTGGGCTTGAATGATCACTATCAAAAGATTTGAAAGACATTGGTTTACT	4975
DB	7745	CAAGACCAAGGCGCTGATTTCAAGCTGTGTAAGAACTGTGAACTGGCTT---T	7801
QY	4976	CAAGTTTACCAATGATGTTTATTTACATGATATGAAGAGGATCAGTATATAT	5035
DB	7802	GAGGCTTAAAGGCTGATACCTCATTTATCTTTTCAAAAGAGGATCTTATGTT	7861
QY	5036	TTAGGTTTATATGATGATATTTCTTATGTTGAAAGTTCAAAAAGTTATGATAT	5095
DB	7862	GTTTGATATATGATGATATATATGATCTAGCTCTACAGAAAGGCAACTACAGCA	7921
QY	5096	TTTGATGATCAATTAGAGATCTTTTGAATTAAGTTGATGAATTCAAATAT	5155
DB	7922	CTTCGAAAGATCTAAACAGAGTTCGACTTAAGATTGGAGACCTCACTACTTC	7981
QY	5156	CTTGATTTGAATTCGTAAACCGAATCGTTATATTTATCTCAAGAAAAATTTCTC	5215
DB	7982	CTTGAAATTGAGTAACTAAAGTTTCCATGCGCTTATCTTGACTCAAGAGATATGCA	8041
QY	5216	AAGAAATTAATTAAGCAATTTCA	5238
DB	8042	AATGATCTGCTAAAGAGATTAA	8064

RESULT 6  
US-10-298-122-3/c  
Sequence 3, Application US/10298122  
Publication No. US2003022121A1  
GENERAL INFORMATION:  
APPLICANT: Giltner, Frederick G  
APPLICANT: Deng, Zhanao  
APPLICANT: Zhang, Hongbin  
TITLE OF INVENTION: CITRUS TRISTEZA VIRUS AND METHODS OF USE  
FILE REFERENCE: 5853-220  
CURRENT APPLICATION NUMBER: US/10/298,122  
CURRENT FILING DATE: 2002-11-15  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 3

LENGTH: 78333  
TYPE: DNA  
ORGANISM: artificial  
FEATURE:  
OTHER INFORMATION: Poncirus trifoliata and citrus grandis hybrid  
US-10-298-122-3

Query Match 1.2%; Score 79.6; DB 15; Length 78333;  
Best Local Similarity 46.1%; Pred. No. 0.00011;  
Matches 385; Conservative 0; Mismatches 439; Indels 12; Gaps 3;

4462 ATTAATGATCCGGATGCTAAAGTCAATGAAATGCTGAATGAGAAATTTGATCAAAAG 4541  
44529 AAAATGATGAATCTTGATTTAGCAATGCAAGAGGAGTGAATCAATTTGAGAAATA 44470  
44542 ATGTTACGAAGAAGTCCCAATTCACCGGTGGAAGCCATATCTATGGGTGGGTAC 4601  
44469 ATGCTGGAGAGTATGCTTCTATCCGGAATATTAATCATTAAGTAAATGGGTAT 44410  
4602 ATACTGAGAAATGATTTCTCTCAAGGTGTTGCG--AAATCAGTTGTTGTC 4658  
44409 TTAGAAACAAATGATGAATCCGGTGTGTTGTAAGAAATAAGCTAGATTAGTGCCTC 44350  
4659 ATGCAACAGCAAAAGGAAATTTGATGACCTTTAGTGTAGTTCACTGTGA 4718  
44349 AAGGTTACCAACCAAGAGGAAATTTGATGAAACCTTTG-----CACCTGAG 44296  
4719 TAGATCTGTGCTAATGATTAATGACATTAATGATGTTGATGAAATGACAAATTC 4778  
44295 CAAGTTGGAATTCATGATGATGTTGATGATGATGATGATGATGATGATGATGATGAT 44236  
4779 AACATTGAGCTGAGTGGCGGTATCTAAATGCTCTATTAATCAATCAATTT 4838  
44235 ATCAATGATGATGCAAAAGTCTTTCTTAATGATGATGATGATGATGATGATGATGAT 44176  
4839 ATGCTTTCCTCCTAATCAGTACCTTTGAAGAAACCAATGTTGTTGTTGAAGCTT 4898  
44175 AACAACTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 44116  
4899 CTGCTATGAGTTAAACAGTGGGTTGATGATGATGATGATGATGATGATGATGATGATGAT 4958  
44115 CTTTATGATGTTAAACAGCACTAGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 44056  
4959 AAGATATGTTTAACTCAAGTTTACCAATGATGATGATGATGATGATGATGATGATGATGAT 5018  
44055 TGGATTAATGATTTTC--AATGGAAAGCGGACACACTCTTTTGTGTTAGCATTAAGA 43939  
5019 AGGATCAGTAT 5078  
43998 ATCAAGACATCTTATGTTCAATATATATATATATATATATATATATATATATATATAT 43939  
5079 AAAAGTATGATATTTTGTGATCAATGAGATCAATTTGAACTTAAATGTTT 5138  
43938 AGTTGTTGTAAGATTTTTCATCATGATGATGATGATGATGATGATGATGATGATGATGAT 43879  
5139 GTGAATATCAAAATATCTGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5198  
43878 GAGATTTAGATTTCTTGTGACTTCAATCAAAAGGAGGAGGATTTTCTATA 43819  
5199 CTCAGAAAAATTTCTCAAGAAATTTCTTAAGATTTCAAACTGATGATCATATATGAGA 5258  
43818 ACCAAGCCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 43759  
5259 AAAACATCCCGGATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5314  
43758 AAGGATCTCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 43703

RESULT 7  
US-10-424-599-118696  
; Sequence 118696, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:

APPLICANT: La Rosa Thomas J  
APPLICANT: Kovalic David K  
APPLICANT: Zhou Yihua  
APPLICANT: Cao Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 118696  
LENGTH: 560  
TYPE: DNA  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_78191C.1  
US-10-424-599-118696

Query Match 1.1%; Score 70.8; DB 12; Length 560;  
Best Local Similarity 51.7%; Pred. No. 0.00057;  
Matches 187; Conservative 0; Mismatches 172; Indels 3; Gaps 1;

4868 AAGAAACCATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 4927  
45 AAGCCATCATGCTCTTAACTGAAAAAGGTTTATATGATTAATGATTAATGATTAATGATTAATGAT 104  
4928 GAATGATACATATATCAATCAAAAGATTTGAAAGATTTGTTTACTCAAGTTTACAC 4987  
105 GCTTGTATGAACTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 161  
4988 AATGATGTTTATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5047  
162 GTTATCTACCTTCTTTTATTAAGAAATGATGATGATGATGATGATGATGATGATGATGATGAT 221  
5048 GTTATGATATTTCTTATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 5107  
222 GTTATGATATATATTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 281  
5108 TTGAGATCATTTTGAATTAAGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 5167  
282 ATGCAAAAGAAATTTGAAATGTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 341  
5168 TTTCGTAACCAAGATCTGTTATATTTATCTCAAGAAATTTCTCAAGAAATTTACTT 5227  
342 ATCAAGCAAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 401  
5228 AA 5229  
402 CA 403

RESULT 8  
US-10-424-599-108338/C  
; Sequence 108338, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
APPLICANT: La Rosa Thomas J  
APPLICANT: Kovalic David K  
APPLICANT: Zhou Yihua  
APPLICANT: Cao Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 108338  
LENGTH: 2760  
TYPE: DNA  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_68844C.1  
US-10-424-599-108338

Query Match 1.1%; Score 70.6; DB 12; Length 2760;  
Best Local Similarity 45.3%; Pred. No. 0.0013;  
Matches 367; Conservative 0; Mismatches 434; Indels 6; Gaps 3;

4496 TGGATTAAGTCATGTAATGCTGTAAGTGAAGTATTTGATCAAGTATGTTTGAAGA 4555  
1503 TGGATCAATGCGATGCAAGAAAGTAAACCAATTTGAAAGAAACATGATGGAATTA 1444  
4556 GTTCCATTTCCACCGGTGAGACCTATATCTATGCGTGGTGGTACCTGAGAAAT 4615  
1443 GTAGAAAAAAGCTGAAATTAATCTGATTAAGAAACAAATGGGTTTTTAATAATTA 1384  
4616 GATTCCTCAAGAGGTGTTGTTGGAATTCACGTTGTGTGTCATGCGACACAGAAAAG 4675  
1383 GATGAACATGTAATTAATTTAGAAATTAAGCCAGGTTAGT---AGAAAGAGGTAAAT 1327  
4676 GAAAAATGATTAAGACCTTTTATGTTAGTTCACCTGTTATAGATCTTGACTATA 4735  
1326 CAAGAGAGGAAATAGACTATGAAGAAACATATGCTCTGTGCAAGATTAGAGCCATT 1267  
4736 AGATTATGACAAATATAGTGTGATTAAGAAATGACAAATTCACATTTAGACGTGAG 4795  
1286 AGAATGCTTTGGCATATGATCCATATGAACTTTAACTTTATCAATGATGTTAG 1207  
4796 TGGCGATCTTAAATGCTCTTATTAATCAATTCATCAATCAATTTATGTTCTCTTAA 4855  
1206 AGTGCCTTTTAAATGCTTATTAATCAAGAGGATATGTTGAACAAACCCCTGTTT 1147  
4856 TCAGTACTTTGAAGAAAAACCATTTGTTGTTATGAAAGTCTGCTATGAGTTAAA 4915  
1146 GAAATTTCTGATTAACCAACCATGTTTATTAATTAACAAAGGCTTTTATGTTGAAA 1087  
4916 CAGTCGGTGTGATGATGATCACTATCAAAAGATTTGGAAGCATTTGTTTACT 4975  
1086 CAAGCCCTAGGAGATGATGACATTAAGTATTTCTCTGTAAGAAAGATTTCTC 1027  
4976 CAAGTTTACCAATGATGTTTATTTTCACTTAATTAATGAAGGATCAATTAAT 5035  
1026 AGAGTAAAGTGAATACCACTTA-TTCATTAAGAGGATATATATTTTGTGGT 968  
5036 TTAGGTTATGTTGATGATATCTTATGTTGGAATTCACAAAGTATTTGTAAT 5095  
967 TCA-AATATATGTTGATGATATTTTGAATCCATATGATTTATGTCAGAGAG 910  
5096 TTTGTGATCAATGAGATCAATTTGAACTTAAAGTGTGGTGAATATCAATTAAT 5155  
909 TTTTCCCTTGATGACAAAGTGAATTTGAATGCAATGATGGAGAACTAAAGTCTTT 850  
5156 CTGGTATGTAATTTGTAAGAAACGATCTGTTATTTTATCTCAAGAAATTTCTC 5215  
849 CTGGATTAACAATACCAACTCAAGAGGTTATTCATCAATCAATCAATCTGC 790  
5216 AAGAAATTAATTAAGATTTGAACTAGATGACTCATATGGAAGAAACATACCTGAT 5275  
789 AAGGATTAATCAAAAGTTGGGATGATGTCACAAACACATGCTACACGATAGC 720  
5276 CCGATGACAAATATGAAGAGTTGCA 5302  
729 ACTAATGTTACTTAATGAAGATGAA 703

RESULT 9  
US-10-312-841-1/c

; Sequence 1, Application US/10312841  
; Publication No. US20030186277A1  
; GENERAL INFORMATION:  
; APPLICANT: Epigenomics AG  
; TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC  
; FILE REFERENCE: E01/1208/WO  
; CURRENT APPLICATION NUMBER: US/10/312,841  
; CURRENT FILING DATE: 2002-12-30  
; NUMBER OF SEQ ID NOS: 2

; SEQ ID NO 1  
; LENGTH: 3673778  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
; NAME/KEY: unsure  
; LOCATION: (3294164)  
US-10-312-841-1

Query Match 1.1%; Score 68.4; DB 14; Length 3673778;  
Best Local Similarity 59.1%; Pred. No. 0.12;  
Matches 117; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

1161 AACAAATGAACTTAATCTTATGTTAGTTTACACAGACCGAGAAACCAAGAAAA 1240  
2552950 AACAACTCTAACCTTTCTTCTCTAATTAATTAATTAATTAATTAATTAATTA 2552891  
1241 CCAGAGGAACTCACTGGAACAACTTCAAAAGAAAGAACTGAAATCAAGAAAGAAAT 1300  
2552890 AAATATCAAAACTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2552831  
1301 AAGAAACATCCAAATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1360  
2552830 AAAAAAAAAAAAAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2552771  
1361 TCATCTGAATGAAAAACA 1378  
2552770 AAAAAAAAAAAAAAAAAA 2552753

RESULT 10

US-10-311-455-1610/c  
; Sequence 1610, Application US/10311455  
; Publication No. US20030143606A1  
; GENERAL INFORMATION:  
; APPLICANT: OLEK, Alexander  
; APPLICANT: PIPERNEROCK, Christian  
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining the Cytosine Methylation of DNA  
; FILE REFERENCE: 5013.1014  
; CURRENT APPLICATION NUMBER: US/10/311,455  
; CURRENT FILING DATE: 2002-12-16  
; PRIOR APPLICATION NUMBER: PCT/EP01/07537  
; PRIOR FILING DATE: 2001-07-02  
; PRIOR APPLICATION NUMBER: DE 10032529.7  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: DE 10043826.1  
; PRIOR FILING DATE: 2000-09-01  
; NUMBER OF SEQ ID NOS: 2424  
; SEQ ID NO 1610  
; LENGTH: 7306  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-311-455-1610

Query Match 1.1%; Score 68.2; DB 14; Length 7306;  
Best Local Similarity 50.1%; Pred. No. 0.0064;  
Matches 169; Conservative 0; Mismatches 168; Indels 0; Gaps 0;

1170 ACATTCAGACCAAAATATGAACTTAATCTTATGTTAGTTTACAGACCAAGAAAC 1229  
6418 AACAACTTAATTAACAAACCACTTATCTTAATTAATTAATTAATTAATTAATTAAT 6359  
1230 CAAAAGAAAGACAGAGAGAACTCACTGAAACATCTCAAAAGAGAACTGAAATCA 1289  
6358 AAAAAAAAAAAAAAAAACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 6299  
1290 GAAAGAGAAATAGAAACATCAAAATCAAGATTAAGATTAAGATTAAGATTAAGATTA 1349

```

TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53223)B
CURRENT APPLICATION NUMBER: US/10/424,599
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 285684
SEQ ID NO 108640
LENGTH: 7628
TYPE: DNA
ORGANISM: Glycine max
FEATURE:
NAME/KEY: unsure
LOCATION: (1)..(7628)
OTHER INFORMATION: unsure at all n locations
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRI3847_69115C.1
US-10-424-599-108640

Query Match 1.0%; Score 65.6; DB 12; Length 7628;
Best Local Similarity 45.1%; Pred. No. 0.021;
Matches 371; Conservative 0; Mismatches 439; Indels 12; Gaps 3;

QY 4484 AATGATCCGGGATGGATTAAGTCAATAGTCACTAGACAGAAATTAGATCAAAAGAT 45433
Db 4263 ACTGATAGTTCTTGATCAATGCTATCCAAAAGAAATTGGACATTCAAAAGGATGCA 43222
QY 4544 GTTACGAGAAGAGTTCATTCCTCCACCGGTGTGAGGCTATATCTATGGTGGGTACAT 46030
Db 4323 GTCTGGGAGGTAGTCTTACGACCCAGGAACTATGATGTGACCAAGATGATCTTC 43828
QY 4604 ACTGAGAAATTTGATTTCTCCAAAGGTGTGTTG---GAAATACGTGTGTGTCCAT 46600
Db 4383 AAGAACAACCAATGAGAAAGGTGTATATCCGAAACAAAGCCCAAGCTGTGTCTCAA 44422
QY 4661 GGCACACAGACAAAAGAAAATTTGAGTTATGACCCCTTTAGTGTAGTTCACTGTTATA 47200
Db 4443 GGCACACCTAGATGGAAGGGTATGCTTGA-----TGAAATTTGGCCCTGTGTCT 44366
QY 4721 GATCTTGTGACTATAGATTATTGACAAATATAGGTTGTGATTTAGGATGACAAATCAA 47800
Db 4497 AGACTTGAGTCCATCAGACTATTACTTGGTGTGAGCTTGCAATCCCAATTCAGCTGTAT 45566
QY 4781 CATTTAGACGTGCGATGGGGATCTTAAATGGCCCTATTTCTCATTCATAATCCAAATTTAT 48400
Db 4557 CAGATGATGTGAAGACCGCGTTTCTGAATGATCCTGATGAAGAACTATGTGAG 46166
QY 4841 GTCTTCTCTCTAATCAGTACCTTTGAGAAAACCAATGTTGGTTATTTGAAACGTTCT 49000
Db 4817 CAGCCAAAGGATTTGACAGATCCCACTATCCAGATCATGATATCAAGCTCAAAAAGGCT 46786
QY 4801 GTCTATGCGTTAAACAGTGGGGTTTGGATGTGATCACTATCAAAAGGTATGGAA 49666
Db 4677 CTCTATGATTTGAAGCAAGCTCCAAAGCTGTGGTATGAAGGCTAACAAAGTCTTACT 47366
QY 4961 GACATTTGTTTACTCAGTTTATACAAATGATGTTATTTCACTTGATATGAAGAG 50200
Db 4737 CAGCAAGGATATGGAAGGAGGAAATGAACAAGCTCTCTTTGTCAAAAGATGTGAA 47866
QY 5021 GGATCAGTATATATTTAGGTTTATATGTTGATGATATTTCTTATGTTGGAAGTTCAAA 50800
Db 4797 AACTGTATGATA--GCACAGATATATGTTGATACATTTGTTTGAAGGATGTGCAAT 48533
QY 5081 AAAGTTATTAATAATTTTGGATCAATTTAGAGATCATTTTGAAGTTAAAGTGTGGT 51400
Db 4854 GAGATGCTTGACATTTTGTCTTCAACAGATGAAATCTGAATTTGATGAGAGCTTGTTGGA 49133
QY 5141 GAAATATCAATATATCTTGTGTATTTGAATTTGCTGTAACCGAATCTGGTTATATTTTATCT 52000
Db 4914 GAGCTGACTATATTTCTGGGACTTCAAGTGAAGAGATGGAGACTCCATATTTCTCTCA 49733
QY 5201 CAAGAAAATTTCTCAAGAAATTAATAAGGATTTCAACCTAGATGACTCATATGGGAAA 52600
Db 4874 CAAGACAGGTATGCAAGAAACATTTGTCAAGAGTTTGGATGGAAATCCAGCATTA 50333

```







This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OW nucleic - nucleic search, using sw model

Run on: April 11, 2004, 13:02:12 ; Search time 9825 Seconds  
(without alignments)  
19531.234 Million cell updates/sec

Title: US-09-430-590E-3

Perfect score: 6426  
Sequence: 1 tgcgtgttcgtgcacattc.....agaagtattatccatca 6426

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
EST:  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estom:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vir:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rtd:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vir1:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	174.6	2.7	1017	29	CNS07BJD AL437951 T7 end of
2	142.4	2.2	905	29	CNS07JAX AL436423 T3 end of
3	127.2	2.0	838	29	CNS07JAX AL436955 T7 end of
4	103.6	1.6	781	12	BI920632 EST540579

5	97.8	1.5	569	9	AI813079	AI813079 20E11 pin
6	96.2	1.5	1035	29	CNS07JAX	AL436647 T7 end of
7	93.6	1.5	832	14	CB970396	CB970396 CAB10003
8	92.8	1.4	766	29	CG964290	CG964290 MBSDX18TF
9	91.4	1.4	866	29	CG937798	CG937798 MBEC46TF
10	91.4	1.4	867	29	CG973377	CG973377 MBEDV30TF
11	91.4	1.4	869	29	CG961676	CG961676 MBEC43TF
12	89.6	1.4	940	29	CG928381	CG928381 MBRT43TF
13	89.4	1.4	803	29	CG821291	CG821291 SOYBA90TF
14	89.2	1.4	926	29	CG943796	CG943796 MBEX57TF
15	88.6	1.4	490	29	AG264218	AG264218 Locust cor
16	88.4	1.4	714	28	CG917427	CG917427 ZMMBBD038
17	87.6	1.4	785	28	BH664738	BH664738 BOHYT34TR
18	87.6	1.4	901	28	CG944168	CG944168 MBEH33TR
19	87	1.4	780	28	BH484021	BH484021 BOGBX72TF
20	86.8	1.4	1200	13	BX437758	BX437758 BX437758
21	86.6	1.3	1098	13	BX377526	BX377526 BX377526
22	86.4	1.3	765	28	BH507520	BH507520 BOGBJ33TF
23	86.2	1.3	688	13	BQ510267	BQ510267 EST617694
24	86.2	1.3	721	28	BH723836	BH723836 BOKMB79TR
25	86.2	1.3	784	28	BH445102	BH445102 BOGSG65TR
26	86	1.3	827	29	CG975352	CG975352 MBEC13TF
27	86	1.3	846	29	CG953096	CG953096 MBEN66TF
28	86	1.3	852	29	CG932551	CG932551 MBEH86TF
29	86	1.3	857	29	CG923622	CG923622 MBENJ21TF
30	86	1.3	875	29	CG932091	CG932091 MBENK14TF
31	86	1.3	875	29	CG971290	CG971290 MBEND57TF
32	86	1.3	890	29	CG934171	CG934171 MBENM25TF
33	86	1.3	923	29	CG941085	CG941085 MBENM86TR
34	85.8	1.3	829	29	BX173672	BX173672 Datto rer
35	85.4	1.3	712	29	CG967635	CG967635 BOIGNS9TR
36	85.2	1.3	844	29	BX139887	BX139887 Datto rer
37	85.2	1.3	914	29	CG956730	CG956730 MBEP89TF
38	85	1.3	827	29	CG961943	CG961943 MBEP94TF
39	85	1.3	905	28	AZ550256	AZ550256 ENTEV58TR
40	84.8	1.3	638	29	AG251630	AG251630 Locust cor
41	84.6	1.3	1130	28	BZ433063	BZ433063 BOKNE19TF
42	84.4	1.3	808	28	BZ505487	BZ505487 BOKNX24TF
43	84.2	1.3	914	29	CG971731	CG971731 MBEP188TF
44	84	1.3	700	28	BH728433	BH728433 BOKNJV64TR
45	84	1.3	945	29	CG936465	CG936465 MBEMH67TR

ALIGNMENTS

RESULT 1	CNS07BJD	1017 bp	DNA	linear	GSS 08-JUL-2001
LOCUS	T7 end of clone BC0AA012E08 of library BC0AA from strain CBS 767 of				
DEFINITION	Debaromyces hansenii, genomic survey sequence.				
ACCESSION	AL437951.1	GI:12221364			
VERSION	AL437951.1				
KEYWORDS	GSS.				
SOURCE	Debaromyces hansenii (anamorph: Candida famata)				
ORGANISM	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;				
REFERENCE	Souci��t, J.L., Aigle, M., Artiguenave, F., Blandin, G.,				
AUTHORS	de Montigny, J., Dujon, B., Durrens, P., Lepingle, A., Lorente, B.,				
	Malpertuy, A., Neuvéglise, C., Ozier-Kalogiropoulos, O., Potier, S.,				
	Saurin, P., Tekala, F., Toffano-Nioche, C., Wesolowski-Jouvet, M.,				
	Winkler, P. and Weissendbach, J.				
	Genomic exploration of the hemiascomycetous yeasts: 1. A set of				
	yeast species for molecular evolution studies				
	FEBS Lett. 487 (1), 3-12 (2000)				
TITLE	JOURNAL	2 (bases 1 to 1017)			
	MEDLINE	Lepingle, A., Casaregola, S., Neuvéglise, C., Bon, E., Nguyen, H.,			
	PUBMED	Artiguenave, F., Winkler, P. and Galliardin, C.			
	AUTHORS				

TITLE Genomic exploration of the hemiascomycetous yeasts: 14.  
JOURNAL Debaryomyces hanseni var. hanseni  
MEDLINE 20584724  
PUBMED 11152889  
REFERENCE 3 (bases 1 to 1017)  
AUTHORS Genoscope.  
TITLE Direct Submission  
JOURNAL Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,  
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex - FRANCE (E-mail :  
segref@genoscope.cns.fr - Web : www.genoscope.cns.fr)  
COMMENT This GSS is part of a random genomic sequencing program of thirteen  
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces  
exiguus, Saccharomyces servazii, Zygosaccharomyces rouxii,  
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces  
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia  
angusta, Debaryomyces hanseni var. hanseni, Pichia sorbitophila,  
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to  
5 kb were prepared and both extremities were sequenced. See  
keywords for description of this sequence and for the sequence of  
the other extremity of this insert.

FEATURES  
source  
Location/Qualifiers  
1..1017  
/organism="Debaryomyces hanseni"  
/mol\_type="genomic DNA"  
/strain="CBS 767"  
/variety="hanseni"  
/db\_xref="taxon:4959"  
/clone="BC0A012E08"  
/clone\_1db="BC0A"  
/note="end : 77"  
misc\_feature  
1..51017  
/note="part of putative transposable element"  
/evidence="not\_experimental"

ORIGIN  
Query Match 2.7%; Score 174.6; DB 29; Length 1017;  
Best Local Similarity 53.2%; Pred. No. 5.3e-21;  
Matches 412; Conservative 1; Mismatches 360; Indels 2; Gaps 2;

4444 CAAAGGATTTACATCGGTCGTCACAAATGACATTAATGATCCGGAGATGATAA 4503  
Db CACTCGGAAATTAATCTGTATATCGAAAGTCGATTAAACAGATTAACAACGATGCA 303  
Qy GTCAATGAATGCTGAATAGAAATTTAGATCAAAAGATGTTACGAAAGTTCAT 4563  
Db GTCAATGACAGCAATTAATGATCTTTAAGAAATAGAGATATACGTTGTGAAAA 363  
Qy TCCACCGGTGGAAGCTATATCTATGCTGGTACATACGAGAAATGATCTCT 4623  
Db TCCTAATAATGTCAAACTATTCCAACTCTGGGTCATACACATTAATTAACATCT 423  
Qy CAAAGGATTTACATCGGTCGTCACAAATGACATTAATGATCCGGAGATGATAA 4683  
Db CAAAGGATTTACATCGGTCGTCACAAATGACATTAATGATCCGGAGATGATAA 483  
Qy GGATTAATGACCTTTTATGATGTTAGTTCACCTGTTATGATCTTGTAATTAAT 4743  
Db ACATGATGATCTCGGAGGTCATCTCGTATGATTAATCAATTAATGATGTTACT 543  
Qy GACATATATAGTGTGATTAAGAAATGACATTAATGATGATGATGATGATGAT 4803  
Db TACAGCGATACAGT-TGAATATGATGATGATGATGATGATGATGATGATGAT 602  
Qy TCTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4863  
Db TTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 662  
Qy TTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4923  
Db TTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 722  
Qy TTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4983

Db GTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 782  
Qy 4984 ACACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5042  
Db 783 TGCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 842  
Qy 5043 TATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5102  
Db 843 TATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 902  
Qy 5103 ATCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5162  
Db 903 TTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 962  
Qy 5163 TTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5217  
Db 963 TTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1017

RESULT 2  
CNS07ACX 905 bp DNA linear GSS 08-JUL-2001  
LOCUS T3 end of clone BC0A003C05 of library BC0A from strain CBS 767 of  
DEFINITION Debaryomyces hanseni, genomic survey sequence.  
ACCESSION AL36423  
VERSION AL36423.1 GI:12219836  
KEYWORDS GSS.  
SOURCE Debaryomyces hanseni (anamorph: Candida famata)  
ORGANISM Debaryomyces hanseni  
1 (bases 1 to 905)  
REFERENCE Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,  
Bojicic-Pukharec,M., Bon,E., Brothier,P., Casaregola,S.,  
de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,  
Malpertuy,A., Neuvéglise,C., Ozier-Kalogreopoulos,O., Potier,S.,  
Saurin,M., Tekala,F., Toffano-Nicote,C., Wesolowski-Louvel,M.,  
Wincker,P. and Weissenbach,J.  
TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of  
Yeast species for molecular evolution studies  
JOURNAL FEMS Lett. 487 (1), 3-12 (2000)  
MEDLINE 20584724  
PUBMED 11152876  
REFERENCE 2 (bases 1 to 905)  
AUTHORS Lepingle,A., Casaregola,S., Neuvéglise,C., Bon,E., Nguyen,H.,  
Artiguenave,F., Wincker,P. and Galliardin,C.  
TITLE Genomic exploration of the hemiascomycetous yeasts: 14.  
JOURNAL Debaryomyces hanseni var. hanseni  
MEDLINE 20584724  
PUBMED 11152889  
REFERENCE 3 (bases 1 to 905)  
AUTHORS Genoscope.  
TITLE Direct Submission  
JOURNAL Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,  
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex - FRANCE (E-mail :  
segref@genoscope.cns.fr - Web : www.genoscope.cns.fr)  
COMMENT This GSS is part of a random genomic sequencing program of thirteen  
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces  
exiguus, Saccharomyces servazii, Zygosaccharomyces rouxii,  
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces  
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia  
angusta, Debaryomyces hanseni var. hanseni, Pichia sorbitophila,  
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to  
5 kb were prepared and both extremities were sequenced. See  
keywords for description of this sequence and for the sequence of  
the other extremity of this insert.

FEATURES  
source  
Location/Qualifiers  
1..905  
/organism="Debaryomyces hanseni"  
/mol\_type="genomic DNA"  
/strain="CBS 767"

```

/misc_feature
/variety="hansenii"
/db_xref="taxon:4959"
/clone="BCOA003C05"
/clone_1b="BCOA"
/note="end : 73"
<1..>905
/note="part of putative transposable element"
/evidence=not_experimental

```

```

Query Match 2.2%; Score 142.4; DB 29; Length 905;
Best Local Similarity 51.0%; Pred. No. 2.9e-15;
Matches 347; Conservative 5; Mismatches 327; Indels 1; Gaps 1;

```

```

4444 CAAGAGGATTTACATGCGGTGCTCAACAAATGACATATATATCCGGATGATAAA 4503
227 CACTCGGAAATTAATCTATATCGAAAAGTCGATTAAACATTAACAACTGGAGCA 286
4504 GTCAATGATGCTGAACCTAGAGAAATTTAGATCAAAAAGTTTACAGAAAGTCCCAT 4563
287 GTCAATACAGCGCAATAGATCTTTTACAGAAATATAGGATATATACGGTTGGAAMA 346
4564 TCCACCGGTGTAAGCTTATATATGCGGTGATCATCTAGAAAATTTGATTTCT 4623
347 TCCATAAATGTCACAACTTATCCAACTTGGTTCATACACATAAATTTDCAATCT 406
4624 CAAGAGTGTGTTGGAATCAGCTGTGTTGTCATGCGCAACAGAAAAGAAATTT 4683
407 CAAGAAGTTGATTAATACACGTTGGTGTACAGGGCTTTAGCAAAATTCGAATGA 466
4684 GGATTTATACCTTTTATGTTAGTTTACCTGTTATATGATCTTTGTGCTATAGATTAT 4743
467 ACACATATATACCTCGAAGGTCATCTCCGTGATGATTTATCCATAATTCGTTACT 526
4744 GACATATATAGTTGATGATAGAGATGACATTCATTTAGCGTCGAGTGGCGTA 4803
527 TACACCGATATAGT-TGAATATGAAATGCCCATACATCATTTATATCTCCGATA 585
4804 TCTAATGCTCTTATCTACTATTCATTCATTCATTAATGCTCTTCTCTTAATCAGTACC 4863
586 TTTACATGACATATCGACATATGAGAAATCCATATTTGTAAACACACCTGATGAA 645
4864 TTTGAGAAAACCATTTGTTGATTTGAACGTTGCTATGCGGTTAAACAGTGGG 4923
646 TATTGATTTCTGTAATGTTGCAATTAACAAATCTGTTTATGGAAGACAGCAGG 705
4924 TTTGGAATGATACACATATCAAAAGATTTGGAAGACATTTGTTTACTCAAGTTT 4983
706 GTATATGATGATCAATGATATCAAGTTCTTATGATCTTAAATTCGAACCTGATAC 765
4984 ACACATATGATGTTTATTTCAATGAAATATGAAGAGGATCAGTAATATTTAGTTT 5043
766 TGCATTTAGCGGAATGTTTGTAAATATTTTGTGAAATTAAGAGCTCATCTGTTGCACT 825
5044 ATATGTTGATATATTTCTTATGTTGGAAGTTCAAAAAGTTTATATTTTGTGA 5103
826 ATATGTTGATATATTTTACTCTGCTATATATTAATTTCAATTTTCAACT 885
5104 TCAATTGAGAGATCAATTTTG 5123
886 TGAACCTCGTAAACATTTTCG 905

```

```

RESULT 3
LOCUS CDS07AP 838 bp DNA linear GSS 08-JUL-2001
DEFINITION T7 end of clone BCOA006C12 of library BCOA from strain CBS 767 of
Debaromyces hansenii, genomic survey sequence.
ACCESSION AL436955
VERSION AL436955.1 GI:12220368
KEYWORDS GSS
SOURCE Debaromyces hansenii (anamorph: Candida famata)
ORGANISM Debaromyces hansenii

```

## REFERENCE

Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Debaromyces.

## AUTHORS

Souchet, J.-L., Aigle, M., Artiguenave, F., Blandin, G., Bolotin-Fukuhara, M., Bon, E., Brotier, P., Casaregola, S., de-Montigny, J., Dujon, B., Durieux, P., Lepingle, A., Llorente, B., Malpertuy, A., Neveu, G., Olier-Kalogeropoulos, O., Portier, S., Sautin, M., Tekala, F., Toffano-Nicolas, C., Wesolowski-Louvel, M., Wincker, P. and Weissbach, J.

## TITLE

Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies

## JOURNAL

FEMS Lett. 487 (1), 3-12 (2000)

## MEDLINE

20584711

## PUBMED

1152876

## AUTHORS

Lepingle, A., Casaregola, S., Neveu, G., Bon, E., Nguyen, H., Artiguenave, F., Wincker, P. and Galliard, C.

## TITLE

Genomic exploration of the hemiascomycetous yeasts: 14. Debaromyces hansenii var. hansenii

## JOURNAL

FEMS Lett. 487 (1), 82-86 (2000)

## MEDLINE

20584724

## PUBMED

1152889

## AUTHORS

3 (bases 1 to 838)

## REFERENCE

Genoscope.

## COMMENT

Submitted (08-SEP-2000) Genoscope - Centre National de Séquençage, 2 rue Gaston Crémieux, CP 5706, 91057 Evry cedex - FRANCE (E-mail : segre@genoscope.cns.fr - Web : www.genoscope.cns.fr)

This GSS is part of a random genomic sequencing program of thirteen yeast species: *Saccharomyces bayanus* var. *uvarum*, *Saccharomyces exiguus*, *Saccharomyces servazzii*, *Zygosaccharomyces rouxi*, *Saccharomyces kluyveri*, *Kluyveromyces thermotolerans*, *Kluyveromyces lacticus* var. *lacticus*, *Kluyveromyces marxianus* var. *marxianus*, *Pichia angusta*, *Debaromyces hansenii* var. *hansenii*, *Pichia sorbicola*, *Candida tropicalis* and *Yarrowia lipolytica*. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.

## FEATURES

Location/Qualifiers

## source

1..838

/organism="Debaromyces hansenii"

/mol\_type="genomic DNA"

/strain="CBS 767"

/variety="hansenii"

/db\_xref="taxon:4959"

/clone="BCOA006C12"

/clone\_1b="BCOA"

/note="end : 77"

<1..>838

/note="part of putative transposable element"

/evidence=not\_experimental

## ORIGIN

Query Match 2.0%; Score 127.2; DB 29; Length 838;  
Best Local Similarity 49.8%; Pred. No. 1.5e-12;  
Matches 308; Conservative 12; Mismatches 298; Indels 1; Gaps 1;

```

4444 CAAGAGTATTTACATGCGGTGCTCAACAAATGACATATATATCCGGATGATAAA 4503
218 CACTCGGAAATTAATCTATATCGAAAAGTCGATTAAACATTAACAACTGGAGCA 277
4504 GTCAATGATGCTGAACCTAGAGAAATTTAGATCAAAAAGTTTACAGAAAGTCCCAT 4563
278 GTCAATACAGCGCAATAGATCTTTTAAAGAAATATAGATATATAGGTTGTAAAT 337
338 TCTATAMATGTCACAACTATTCACACTACCTGGTTATACACATAMATTAGACACT 397
4624 CAAGAGTGTGTTGGAATCAGCTGTGTTGTCATGCGCAACAGAAAAGAAATTT 4683
398 CAAGAAGTTGATTAATATCACTGTTGTTGATACAGGCTTTAGCAAAATTCGAATGA 457

```

QY 4684 GGATTATGACCTTTAGTGTAGTTCACCTGTTATAGATCTTGATGACTAATGATTATT 4743  
 Db 458 ACATATGATACCTCCGAGAGTGTACATCTCTGATGATGACTTATCATATATTCGTTACT 517  
 QY 4744 GACAAATATAGTGTGATGATTAAGAAATGAAATTCAACTTTGAGGTGAGTGGCGGTA 4803  
 Db 518 TACAGGCGATGAGT-TGAAATATGACTGGCCGATACATCATCTGAATATATCTCCGCAATA 576  
 QY 4804 TCTAATGCTCTATATACATTCATTCAAATTTATGTTCTTCTCTTAAATCACTACC 4863  
 Db 577 TTTCCCGCCGATATGACATATCASNATATTCATATTCATACCCCAACCTGGCTGAC 636  
 QY 4864 TTGAGAAAAACCATGTTGTTGTTATTGAAGCTTCTGCTATGAGTTAAACAGTGGG 4923  
 Db 637 TATGATATTCGTAATATTTGTAATTTGTAATTTGTAATTTGTAATTTGTAATTTGTAATTT 696  
 QY 4924 TTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4983  
 Db 697 GTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 756  
 QY 4984 ACACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5043  
 Db 757 TGCATATGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 816  
 QY 5044 ATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5062  
 Db 817 ATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 835

RESULT 4  
 B1920632 781 bp mRNA linear EST 10-MAR-2003  
 LOCUS B1920632 potato microtubers, in vitro-grown Solanum tuberosum cDNA  
 DEFINITION clone cSTE26617 5' end, mRNA sequence.

ACCESSION B1920632  
 VERSION B1920632  
 KEYWORDS EST.  
 SOURCE Solanum tuberosum (potato)  
 ORGANISM Solanum tuberosum

REFERENCE  
 AUTHORS Van der Hoeven, R., Bezzerides, J., Bachem, C., Visser, R., Karamcheva, S.A., Tsai, J., Van Aken, S., Uteckback, T., Chiemingo, A., Bougri, O., Buell, C.R., Rinning, C., Tanksley, S. and Baker, B.  
 TITLE Generation of ESTs from in vitro grown microtubers (2001b)  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Robin Buell

The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Email: potato-array@ligr.org  
 This clone can be obtained from the University of Arizona Genomics  
 Institute. Orders can be made through URL:  
 http://genome.arizona.edu/orders/  
 Seq primer: T3.

FEATURES  
 source  
 Location/Qualifiers  
 1..781  
 /organism="Solanum tuberosum"  
 /mol\_type="mRNA"  
 /cultiVar="Bintje"  
 /db\_xref="taxon:4113"  
 /clone="cSTE26617"  
 /tissue\_type="axillary buds of stem explants; growing sink-tubers"  
 /dev\_stage="7, 8 and 10 days"  
 /lab\_host="SOLR"  
 /clone\_lib="potato microtubers, in vitro-grown"  
 /note="Vector: pBluescript SK(-); Site1: EcorI; Site2: XhoI; tissue supplied by Christian Bachem and Richard Visser (Department of Plant Breeding, Wageningen University, The Netherlands); sequencing by the institute for Genomic Research. The cDNA libraries will attempt to

capture the induction and initiation/initial growth of the tuber in an in vitro system as described in Bachem et al. (Plant Journal, 1996). Small microtubers develop from axillary buds attached to stem explants when placed on a high sucrose medium (10%). Visible morphological changes occur synchronously at day five in the axillary buds. The first library, cSTA (1-20) consists of axillary buds harvested on days 1-3. This targets those genes involved in induction of the microtubers. The following libraries, cSTA (21-40) and cSTA (41-60), capture genes involved in tuber initiation and outgrowth. This library is noted as p3 in Tanksley lab notebooks."

## ORIGIN

Query Match 1.6%; Score 103.6; DB 12; Length 781;  
 Best Local Similarity 48.0%; Pred. No. 2.3e-08;  
 Matches 360; Conservative 0; Mismatches 384; Indels 6; Gaps 2;

QY 4486 TGATCCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4545  
 Db 15 TCATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 74  
 QY 4546 TTAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4605  
 Db 75 TTGGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 134  
 QY 4606 TGAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4665  
 Db 135 TATCAAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 191  
 QY 4666 CAGACAAAGAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4725  
 Db 192 AGGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 251  
 QY 4726 TGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4785  
 Db 252 TGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 311  
 QY 4786 AGACGTCAGTGGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4845  
 Db 312 GGACATTAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 371  
 QY 4846 TCCTCTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4902  
 Db 372 ACCGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 431  
 QY 4903 CTATGCGTTAAACAGTGGCGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4962  
 Db 432 CTATGCTTAACAGTGGCGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 491  
 QY 4963 CATTGTTTACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5022  
 Db 492 GTTGGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 551  
 QY 5023 ATCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5082  
 Db 552 TCAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 611  
 QY 5083 AGTTATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5142  
 Db 612 TGGTATCAACGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 671  
 QY 5143 AATATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5202  
 Db 672 ATTGAAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 731  
 QY 5203 AGAAATTTTCAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5232  
 Db 732 ACGCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 761

RESULT 5  
 A1813079

LOCUS	AI813079	569 bp	mRNA	linear	EST 08-JUL-1998
DEFINITION	20E11 Pine Lambda zap	xylem library	Pinus taeda cDNA, mRNA		
ACCESSION	AI813079				
VERSION	AI813079.1	GI:5424175			
KEYWORDS	EST.				
SOURCE	Pinus taeda (loblolly pine)				
ORGANISM	Pinus taeda				
REFERENCE	Eukaryota, Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.				
AUTHORS	Whetten,R.W., Kinlaw,C.S., Retzel,E. and Sederoff,R.R.				
TITLE	The Pine Gene Discovery Project				
JOURNAL	Unpublished (1999)				
COMMENT	Contact: Ross Whetten Forest Biotechnology Group North Carolina State University Dept. of Forestry, NC State University, 6113 Jordan Hall, Raleigh, NC, 27695-8008 Tel: 919-515-7800 Fax: 919-515-7801 Email: rosswhetten@unity.ncsu.edu Seq primer: T3.				
FEATURES					
source	location/Qualifiers				
	1..569				
	/organism="Pinus taeda"				
	/mol_type="mRNA"				
	/db_xref="taxon:3352"				
	/tissue_type="differentiating xylem"				
	/clone_id="Pine Lambda Zap xylem library"				
	/note="Vector: Lambda Zap; Site 1: EcoRI; Site 2: XhoI; differentiating xylem was collected from the main stem of a 35-year old loblolly pine tree harvested during the growing season. RNA isolation and library preparation followed the methods of Allona et al., PNAS 95:6692-8, 1998"				
ORIGIN					
Query Match	1.5%;	Score 97.8;	DB 9;	Length 569;	
Best Local Similarity	48.5%;	Pred. No. 2.6e-07;			
Matches 270;	Conservative 0;	Mismatches 287;	Indels 0;	Gaps 0;	
Oy	4730	ACTATAATTATTCAGCAATAATACGTTGGATAGGAATGACATTCACATTTAAG	47839		
Db	13	ACAATTAAATTAATCTGATTCCTTACCCACAAAACATAATGGACATCACTCAATGGAT	72		
Oy	4790	GTCGAGTCGGGGATCTAAATGCGCTCTTACTCATTCATAATCCAACTTATGCTCTTCT	48449		
Db	73	GTCAAATCTGCTTTCTTAATGGTGATTTAAAGGAAGAAGTTAATTTGGTTCACCTGAA	132		
Oy	4850	CCTAATACGATACCTTTGAAGAAAAACATTTGTTGTTATTTGAAACGTTCTGTCTATGG	49039		
Db	133	GGGTTTGTCAAAACAGGGAACAGAGCATGTGGTTGACAGTTAAGAAAGCATTTGATGT	192		
Oy	4910	TTAAACACAGTCGGGTTTGAATGATCACACTATCAAAAGATTTGGAAGACATTTG	49659		
Db	193	TTGAACAGGACCAAGCATGTGATGATTAAGATGATGATACCTCTTCTTCAGAAATGA	252		
Oy	4970	TTTACTCAAGTTTACACATGATGTTTATTTCACTGAATTAATGAAGGGATCAGTA	50299		
Db	253	TTTGTGAAGAACAAAAAGATCCTAACCTGTATGTTAAGAAAGATGAAAAAGAAATGTT	312		
Oy	5030	ATATATTTAGGTTTATATGTTGATGATATTTCTTAATGTTGAAGTTCAAAAAAGTAT	50899		
Db	313	GCTTAAATATCTTTGTATGTGATGATCTCATTTATAAGGTAGTCTCAAGTTAAT	372		
Oy	5090	GATAATTTTGTGATCAATTGAGAGATCATTGTTGAAGTTAAAGTTTGTGAAATATCA	51499		
Db	373	GAAAGAAATTAAGTCACTGTTCTCAGAGAAATTTGAATGAAGAAGATTTGGGGAATGAT	432		
Oy	5150	AATATCTGTGATTTGAATTTGCTAAACGAACTCGTTATATTTTATCTCAGAAATA	52099		
Db	433	TACTGTTTAAAGATTGAGTATGAGAGAACCTGTGTTAAACCTTGATCACTCAAGCAAA	492		

Oy		5210	TTTCTCAGAAATTAAGTATTGCAACCTAATGACTCATTTGGAAAAACAATGCC	526259
Dd		493	TATCAAAGAAGATTCCTTAATAAGTTCAATATGTGCAAGATTCCAAAGCTATGTCAATCCC	552
Oy		5270	TGGATTCCGAATGACAA	5286
Dd		553	TTAGAACAGATGCMAA	569

  

RESULT 6	CNS07AOP/c	CNS07AOP	1035 bp	DNA	linear	GSS 08-JUN-2001
LOCUS	DEFINITION	T7 end of clone BCOA005F10 of library BCOA from strain CBS 767 of Debaryomyces hansenii, genomic survey sequence.				
ACCESSION	VERSION	AJ436847				
KEYWORDS	SOURCE	AJ436847.1 GI:12220260 GSS.				
ORGANISM	Debaryomyces hansenii (anamorph: Candida famata)					
REFERENCE	AUTHORS	Debaryomyces hansenii Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Debaryomycetales; Saccharomycetaceae; Debaryomyces. 1 (bases 1 to 1035) Socciat,J.L., Aigle,M., Attienave,F., Blandin,G., Boitron-Fuhkharz,M., Bon,E., Brotier,P., Casaregola,S., de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B., Maupertuy,A., Neuvéglise,C., Ozer-Kalogeropoulos,O., Potier,S., Saurin,W., Tekala,F., Toffano-Nioche,C., Weslowski-Louvel,M., Wincker,P. and Weissenbach,J. Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies FEMS Lett. 487 (1), 3-12 (2000)				
JOURNAL	MEDLINE	20584711				
PUBMED	REFERENCE	11152876				
AUTHORS	TITLE	2 (bases 1 to 1035) Lepingle,A., Casaregola,S., Neuvéglise,C., Bon,E., Nguyen,H., Attienave,F., Wincker,P. and Galliardin,C. Genomic exploration of the hemiascomycetous yeasts: 14. Debaryomyces hansenii var. hansenii FEMS Lett. 487 (1), 82-86 (2000)				
JOURNAL	MEDLINE	20584724				
PUBMED	REFERENCE	11152889				
AUTHORS	TITLE	3 (bases 1 to 1035) Genoscope. Direct Submission Submitted (08-SEP-2000) Genoscope - Centre National de Séquençage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex - FRANCE (E-mail : segre@genoscope.cns.fr - Web : www.genoscope.cns.fr) This GSS is part of a random genomic sequencing program of thirteen yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces exiguus, Saccharomyces servazzii, Zygosaccharomyces Rouxii, Saccharomyces kluyveri, Kluyveromyces thermocolerans, Kluyveromyces lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila, Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 Kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert. Location/Qualifiers 1..1035 /organism="Debaryomyces hansenii!" /mol_type="genomic DNA" /strain="CBS 767" /variety="hansenii!" /db_xref="taxon:4959" /clone="BCOA005F10" /clone_lib="BCOA" /note="end : T7" <1..>1035 /note="part of putative transposable element" /evidence=not_experimental				

  

FEATURES

source

misc\_feature

Origin





## COMMENT

Other GSSs: MBECD18TRC  
 TIGR  
 Contact: Chris Town  
 9712 Medical Center Drive, Rockville, MD 20850, USA.  
 Tel: 301-838-3523  
 Fax: 301-838-0208  
 Email: cdtown@tigr.org  
 Seq primer: CAGGAAACGCTATGACC  
 Class: BAC ends

## FEATURES

source

1..766  
 /organism="Medicago truncatula"  
 /mol\_type="genomic DNA"  
 /cultivar="genotype A17"  
 /db\_xref="taxon:3880"  
 /clone="35D12"  
 /clone\_1lib="mbh2"  
 /note="Vector: pBelobAC11; Site\_1: HindIII; Site\_2:  
 HindIII; Cook, D.R. and Kim, D.J, unpublished"

## ORIGIN

Query Match 1.4%; Score 92.8; DB 29; Length 766;  
 Best Local Similarity 47.0%; Pred. No. 1.9e-06;  
 Matches 286; Conservative 0; Mismatches 322; Indels 0; Gaps 0;

4725 TTGTGACTATAAGATTATGACAAATAATAGGTTGTGAATTAGAAATGACAAATTCACATT 4784  
 73 TGGATACATTCGAAATCTGTAGCGTTGGCAGTACAAATAAGTGAAATTTGGCATT 132  
 4785 TAGACGTGAGTCGGCGATCTAAATGCTTATTACTCATTCACATTTATGTCT 4844  
 133 TAGATTTAATTCAGCATCTTTAAATGGAATCTGATGAAGAATTTATGTGCTCAC 192  
 4845 TTCCCTTAATTCAGTACCTTTGAAGAAAACCATTTGGTTATTTGAACGTTCTCT 4904  
 193 CTGCTGTTTTTTGTTAAAGGAGGAGAACAGTGATTAACCTTCATTAAGCTTTGT 252  
 4905 ATGGGTTAAACAGTCGGGTTGGAATGCTATCACTATCAAAAGAGTATGGACACA 4964  
 253 ATGGGCTGAAACAGGCCCTAGAGCTGTGACAACAAAATTAATTCACCTTTCTTAATC 312  
 4965 TTGCTTACTCAAGTTTACACATGATGTTTATTTACATTTGAATGAGAGGAT 5024  
 313 AAGATTTAAAGAGATGAAATGATGCACTTTATGTAAGAGCTTTGTGATGTG 372  
 5025 CAGTAATATATTAGTTTATATGTTGATGATTTTATTTAGTTGAAGTTCAAAAAG 5084  
 373 GTTCCTTATTTGCTCTTTGATGTGATGACCTGTAGTAAACAAGCAATATCAACAAG 432  
 5085 TTATTGATATTTTGTGATCAATTTGAGAGTCATTTTGAAGTTAAAGTTTGTGTTAAA 5144  
 433 AAGTACATCAACTTATGAGAGATGAAAGAACAGTTGAGATCTTAGCTTAGGGGAAA 492  
 5145 TATCAAAATTTATCTGTTATGAAATTTGTAACCAATCGTTATTTTATCTCAAG 5204  
 493 TGAACATTTTCTTGCTTGAAGATGATCAATCTGAGAGTGAATTTTTTGTGATCAAG 552  
 5205 AAAAATTTCTCAGAAATTTACTTAAAGATTTCAAACTAGATGATCTCATATGGGAAAAACA 5264  
 553 AAGAGAAATGCTCAGAAATTTTGAAGAAGTTTAAATGGAAGCTGCAAACTGCTCCAA 612  
 5265 TACCGTGATTCGGAATGACAAATATGAAAGGTTGCAATATCTGTAAGAGCTTAATC 5324  
 613 CTCCTTTGTGTGAATTTGAACTCTCAAGGAAGATGAAGCTGATTAATTTGATGCTT 672  
 5325 CAGAGAAT 5332  
 673 CTCATTAT 680

RESULT 9  
 CG937798 866 bp DNA linear GSS 12-DEC-2003  
 LOCUS CG937798

## DEFINITION

MBECA46TF mth2 Medicago truncatula genomic clone 24G19, genomic survey sequence.

ACCESSION  
 VERSION  
 CG937798.1 GI:39809821  
 GSS:  
 KEYWORDS  
 SOURCE  
 ORGANISM

Medicago truncatula (barrel medic)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eustosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;  
 Medicago.

REFERENCE  
 1 (bases 1 to 866)  
 Town, C.D., Shetty, J., Koo, H. and Feldblyum, T.F.  
 Sequencing of BAC ends from Medicago truncatula  
 Unpublished (2003)  
 JOURNAL  
 COMMENT  
 Other GSSs: MBECA46TR

Contact: Chris Town

TIGR  
 9712 Medical Center Drive, Rockville, MD 20850, USA.  
 Tel: 301-838-3523  
 Fax: 301-838-0208

Email: cdtown@tigr.org  
 Seq primer: TGTAAACGACGCCAGT  
 Class: BAC ends.

## FEATURES

source

1..866  
 /organism="Medicago truncatula"  
 /mol\_type="genomic DNA"  
 /cultivar="genotype A17"  
 /db\_xref="taxon:3880"  
 /clone="24G19"  
 /clone\_1lib="mbh2"  
 /note="Vector: pBelobAC11; Site\_1: HindIII; Site\_2:  
 HindIII; Cook, D.R. and Kim, D.J, unpublished"

## ORIGIN

Query Match 1.4%; Score 91.4; DB 29; Length 866;  
 Best Local Similarity 47.5%; Pred. No. 3.4e-06;  
 Matches 272; Conservative 0; Mismatches 301; Indels 0; Gaps 0;

4725 TTGTGACTATAAGATTATGACAAATAATAGGTTGTGAATTAGAAATGACAAATTCACATT 4784  
 73 TGGATACATTCGAAATCTGTAGCGTTGGCAGTACAAATAAGTGAAATTTGGCATT 132  
 4785 TAGACGTGAGTCGGCGATCTAAATGCTTATTACTCATTCACATTTATGTCT 4844  
 133 TAGATTTAATTCAGCATCTTTAAATGGAATCTGATGAAGAATTTATGTGCTCAC 192  
 4845 TTCCCTTAATTCAGTACCTTTGAAGAAAACCATTTGGTTATTTGAACGTTCTCT 4904  
 193 CTGCTGTTTTTTGTTAAAGGAGGAGAACAGTGATTAACCTTCATTAAGCTTTGT 252  
 4905 ATGGGTTAAACAGTCGGGTTGGAATGCTATCACTATCAAAAGAGTATGGACACA 4964  
 253 ATGGGCTGAAACAGGCCCTAGAGCTGTGACAACAAAATTAATTCACCTTTCTTAATC 312  
 4965 TTGCTTACTCAAGTTTACCAATGATGTTTATTTACATTTGAATGAGAGGAT 5024  
 313 AAGGATTTAAAGAGATGAAATGATGCACTTTTATGTAAGAGCTTTGTGATGTG 372  
 5025 CAGTAATATATTAGTTTATATGTTGATGATTTTATTTAGTTGAAGTTCAAAAAG 5084  
 373 GTTCCTTATTTGCTCTTTGATGTGATGACCTCTGATTAACAAGCAATATCAACAAG 432  
 5085 TTATTGATATTTTGTGATCAATTTGAGAGTCATTTTGAAGTTAAAGTTTGTGTTAAA 5144  
 433 AAGTACATCAACTTATGAGAGATGAAAGAACAGTTGAGATGCTTAGCTTAGGGGAAA 492  
 5145 TATCAAAATTTATCTGTTATGAAATTTGTAACCAATCTGTTATTTTATCTCAAG 5204  
 493 TGAACATTTTCTTGCTTGAAGATGATCAATCTGAGAGTGAATTTTTTGTGATCAAG 552  
 5205 AAAAATTTCTCAGAAATTTACTTAAAGATTTCAAACTAGATGATCTCATATGGGAAAAACA 5264

Db 553 AGAAGATGCTCAATGAAGTTTGAAGAGTTAAATGAGAACTGCAATCTGCTCA 612  
QY 5265 TACCTGGATTCCGAATGACAAATATGAAGA 5297  
Db 613 CTCCTTTGGTGGAAATTTGAACCTCTCAAG 645

## RESULT 10

CG973377.

LOCUS 867 bp DNA linear GSS 15-DEC-2003  
DEFINITION MBEV307FC mch2 Medicago truncatula genomic clone 35E12, genomic  
survey sequence.

ACCESSION CG973377  
VERSION CG973377.1 GI:39839156  
KEYWORDS GSS.  
SOURCE Medicago truncatula (barrel medic)  
ORGANISM Medicago truncatula

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;  
Medicago.

1 (bases 1 to 867)

REFERENCE Town, C.D., Shetty, J., Koo, H. and Feldblyum, T.F.  
AUTHORS Sequencing of BAC ends from Medicago truncatula  
JOURNAL Unpublished (2003)  
COMMENT Other GSSs: MBEV307RC  
Contact: Chris Town

TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523  
Fax: 301-838-0208

Email: cdtown@tigr.org

Seq primer: CAGGAACAGCTATGACC

Class: BAC ends.

FEATURES  
source location/Qualifiers

1..867  
/organism="Medicago truncatula"  
/mol\_type="genomic DNA"  
/cultivar="genotype A17"  
/db\_xref="taxon:3880"  
/clone="35E12"  
/clone\_1lb="mch2"  
/note="Vector: pBelobAC11; Site\_1: HindIII; Site\_2:  
HindIII; Cook, D.R. and Kim, D.J. unpublished"

## ORIGIN

Query Match 1.4%; Score 91.4; DB 29; Length 867;  
Best Local Similarity 47.5%; Pred. No. 3.4e-06;  
Matches 272; Conservative 0; Mismatches 301; Indels 0; Gaps 0;  
QY 4725 TTGCACTATAGATTAATGCAATATAGTTGTGATTAGGAATGACAAATTCACATT 4784  
Db 73 TGGATCACTATGCAATCTTGATGCGATGACAAATTAAGTGAATTTGGCATT 132  
QY 4785 TAGACGTGAGTCGGCGATCTTAATGCTCTATTACTCAATTCAAATTTATGTCT 4844  
Db 133 TAGATTTCAATTCAGCAATCTTAATGGAATCTTGATGAAGAAATTTATGTCACAC 192  
QY 4845 TTCTCTCTAATCTGATCTTTGAAGAAAACATTTGTTGTTATTTGAACGTTCTCT 4904  
Db 193 CTGCTGTTTTTTGGTAAAGGAGGAGGAGCAAGGTGTAAAGCTTCATAAAGCTTTGT 252  
QY 4905 ATGGGTTAAACAGTCGGGTTTGGATGCTATCACTATCAAAAGATTTGGAAGACA 4964  
Db 253 ATGGGCTGAAGAACAGCCCTAGAGCTTGTAACAATAATTAATTCACATTTCTTAATC 312  
QY 4965 TTGCTTTTCAAGTTTACACATATGTTTATTTTCACTTTGAATATGAAGAGGAT 5024  
Db 313 AAGGATTCAAAGAGGAGGAATGATCAACTCTTATGTGAAGAGCTTTGGATGGTG 372  
QY 5025 CAGTAATATATTTAGTTTATATGTTATGATCAATCTTATGTTGGAGTTCACAAAAG 5084

Db 373 GTTCTTAATGTCTCTTTGTATGTGATGACCTGCTAGTAACCAATATACACAG 432  
QY 5085 TTATGATTAATTTTGTGATCAATTTGAGATCATTTTGAAGTTAAGTTTGGTGA 5144  
Db 433 AAGTACATCACTATATGAGAGATGAAGAACAGTTTGATGTCTAGCTTAGGGGAAA 492  
QY 5145 TATCAATTAATCTTTGATTTGAATTTGTTAAACCGAATCTGTTATATTTTATCTAAG 5204  
Db 493 TGAATATTTCTTTGGCTTGGAGTGCATCATCTGAGGTGAATTTTTTGAATCAG 552  
QY 5205 AAAATTTCTCAAGAAATTAATTAAGATTTCATAGATGATGATCTCATATGGGAAAAACA 5264  
Db 553 AGAAGATGCTCAATGAAGTTTGAAGAGTTTAAATGGAAGAGTGCAAATCTGCTCA 612  
QY 5265 TACCTGGATTCCGAATGACAAATATGAAGA 5297  
Db 613 CTCCTTTGGTGGAAATTTGAACCTCTCAAG 645

## RESULT 11

CG961676

LOCUS 869 bp DNA linear GSS 15-DEC-2003  
DEFINITION MBECA357F mch2 Medicago truncatula genomic clone 24E21, genomic  
survey sequence.

ACCESSION CG961676  
VERSION CG961676.1 GI:39883571  
KEYWORDS GSS.  
SOURCE Medicago truncatula (barrel medic)  
ORGANISM Medicago truncatula

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;  
Medicago.

1 (bases 1 to 869)

REFERENCE Town, C.D., Shetty, J., Koo, H. and Feldblyum, T.F.  
AUTHORS Sequencing of BAC ends from Medicago truncatula  
JOURNAL Unpublished (2003)  
COMMENT Other GSSs: MBECA357R  
Contact: Chris Town

TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523  
Fax: 301-838-0208

Email: cdtown@tigr.org

Seq primer: TGTAAACGAGCGCCAGT

Class: BAC ends.

FEATURES  
source location/Qualifiers

1..869  
/organism="Medicago truncatula"  
/mol\_type="genomic DNA"  
/cultivar="genotype A17"  
/db\_xref="taxon:3880"  
/clone="24E21"  
/clone\_1lb="mch2"  
/note="Vector: pBelobAC11; Site\_1: HindIII; Site\_2:  
HindIII; Cook, D.R. and Kim, D.J. unpublished"

## ORIGIN

Query Match 1.4%; Score 91.4; DB 29; Length 869;  
Best Local Similarity 47.5%; Pred. No. 3.4e-06;  
Matches 272; Conservative 0; Mismatches 301; Indels 0; Gaps 0;  
QY 4725 TTGCACTATAGATTAATGCAATATAGTTGTGATTAGGAATGACAAATTCACATT 4784  
Db 73 TGGATCACTATGCAATCTTGATGCGATGACAAATTAAGTGAATTTGGCATT 132  
QY 4785 TAGACGTGAGTCGGCGATCTTAATGCTCTATTACTCAATTCAAATTTATGTCT 4844  
Db 133 TAGATTTCAATTCAGCAATCTTAATGGAATCTTGATGAAGAAATTTATGTCACAC 192  
QY 4845 TTCTCTCTAATCTGATCTTTGAAGAAAACATTTGTTGTTATTTGAACGTTCTCT 4904  
Db 193 CTGCTGTTTTTTGGTAAAGGAGGAGGAGCAAGGTGTATTAAGCTTCATAAAGCTTTGT 252

QY 4905 ATGGGTTAAACAGCTGGGTTTGGATGTATCACACTATCAAAAGATTTGGAGACA 4964  
 DB 253 ATGGGCTGAACAGCCCTTGTAGCTGTATCAACAAATTAATATCCACTTTCTTAATC 312  
 QY 4965 TTGGTTTACTCAAGTTTACACATGATGTTTATTTACATTTGAATGAAAGAGGAT 5024  
 DB 313 AAGATTCAAAAGAGGAAATGATGCACTCTTTATGTAAGAGCTTTGGATGGTG 372  
 QY 5025 CAGTAATATATTTAGTTTATATGTATGTATTTATTTATTTGTAAGTTCAAAAAG 5084  
 DB 373 GTTCTTAAATGCTCTTTGTATGTATGTATGACCTGCTAGTAACAGCAATATCAACAG 432  
 QY 5085 TTATGTATATTTTGTGATCAATTTGAGATCAATTTTGAAGTTAAAGTGTGGTGA 5144  
 DB 433 AAGTACATCAACTATGAGAGAGATGAAAAACAGTTTGAAGATGTACGCTTAAAGGGAAA 492  
 QY 5145 TATGAATTTATCTGTATGTAATTTGTAATTTGTAATTTGTAATTTTATTTATCTGAAG 5204  
 DB 493 TGAACATTTTCTTGGCTTGGAGATGATCAATTTGAGATGGAATTTTGTGAATCAAG 552  
 QY 5205 AAAAATTTCTAGAAAATTTACTTAAGATTTCAAACTGATGACTCATATGAGAAAAACA 5264  
 DB 553 AGAAGATGCTCATGAAGTTTGAAGAAGTTAAATGGAAGCTGCAAAATGCTCCAA 612  
 QY 5265 TACCTGTATTCGATGACATGACAAATATGAAAG 5297  
 DB 613 CTCCTTGTGTGTAATTTGAACTCTCAAGG 645

## RESULT 12

LOCUS CG928381 940 bp DNA linear GSS 12-DEC-2003  
 DEFINITION MBE1G3TF mth2 Medicago truncatula genomic clone 61H13, genomic survey sequence.  
 ACCESSION CG928381  
 VERSION CG928381.1 GI:39790015  
 KEYWORDS  
 SOURCE  
 ORGANISM Medicago truncatula (barrel medic)

REFERENCE 1 (bases 1 to 940)  
 AUTHORS Town, C.D., Shetty, J., Koo, H. and Feldblum, T.F.  
 TITLE Sequencing of BAC ends from Medicago truncatula  
 JOURNAL Unpublished (2003)  
 COMMENT Other GSSs: MBE1G43TRB  
 CONTACT: Chris Town  
 TIGR  
 9712 Medical Center Drive, Rockville, MD 20850, USA.  
 Tel.: 301-838-3523  
 Fax: 301-838-0208  
 Email: cdtown@tigr.org  
 Seq primer: TGTAAAACGACGGCCAGT  
 Class: BAC ends.

## FEATURES

LOCATION/Qualifiers  
 1..940  
 /organism="Medicago truncatula"  
 /mol\_type="genomic DNA"  
 /cultivar="genotype A17"  
 /db\_xref="taxon:3880"  
 /clone="61H13"  
 /clone\_id="mth2"  
 /note="Vector: pBELOBAC11; Site 1: HindIII; Site 2: HindIII; Cook, D.R. and Kim, D.U., unpublished"

## ORIGIN

Query Match 1.4%; Score 89.6; DB 29; Length 940;  
 Best Local Similarity 47.5%; Pred. No. 7e-06;  
 Matches 365; Conservative 0; Mismatches 394; Indels 9; Gaps 3;

QY 4483 TAATGATCCGGATGATTAAGTCAATGATGCTGAAGTAAGAAATTTAGATCAAAAAGA 4542  
 DB 118 TAAACATGATGTTGGAGACATGCTATGCAAGTTGACCTTCAAGCTCTTGAGAAAATCG 177  
 QY 4543 TGTTTACGAGAAAGTTCCAAATTTCCCAACGGGTGTAGAGCTTATCTATGTTGGTGACA 4602  
 DB 178 TACTGAAAACCTTGATTTTACACCAAAATGTTAAGCTTATGATGTATGATGATTTA 237  
 QY 4603 TACTGAAAATTTGATTTCTCAAAAGTGTGTTGGAAATCACGTTGTGTCATGG 4662  
 DB 238 CAAAGTCAAAATTTCAATGCT---GATGGCATTATGAGATATAAGCAAGATGTTGGC 294  
 QY 4663 CAACAGCAAAAAGAAAATTTGATTTATGACCTTTTGTGTTAGTTCACTGTTATGA 4722  
 DB 295 CAAAGTTTACATCAAAATTTGAGAGGCTTATGATTTATTAATCATATTTCTCAAGTTCTAA 354  
 QY 4723 TCTTGACTATTAATATTTATGACATTAATAGTTGTGAATTAAGGATGACATTCACA 4782  
 DB 355 ACTCACACAGTCACACTTGTGATTTGCTTTATCTTCATTCATTAATGGCATTTACATCA 414  
 QY 4783 TTTAAGCTGAGTGGCGATCTAATGCTCTAATCTCATTCATCAATTCATTAATGT 4842  
 DB 415 ACTTATATGCAATTAATGCTTTCTTCATG---GTCACTTACAGAAAGATGTATCATGCT 471  
 QY 4843 CTTTCTCTCTAATATGATACCTTTGAGAAAACCATTTGTTGTTATTTGAACGTTCTGT 4902  
 DB 472 GATTCCTCTGGCATCAACCATCAACAAATCAATGTTGCAAGCTTCAAAAGTCTCT 531  
 QY 4903 CTATGGTTTAAACAGTCCGTTTGAATGATGATCACTATCAAAAGATTTGGAAGA 4962  
 DB 532 CTATGACTTAATATGCTGATGAGAAAGATGAGAAATGATGATCTTACTCTTTGGC 591  
 QY 4963 CATTCGTTTACTCAAGTTTACACATGATGCTTTATTTACATGTAATATGAAGAG 5022  
 DB 592 TCAAGCATTAATTTCAAGCTGATGATCATTCCTTGTGTTGCAAGAACTGAA---TC 648  
 QY 5023 ATCAGTAATATTTAGTTTATGTTATGATGATGATTTCTTATGTTGGAAGTTCAAAA 5082  
 DB 649 ATCATTCACAGTCTCTTGTTGATGATGATGATGATGATGATGATGATGATGATGATG 708  
 QY 5083 AGTTATGATTAATTTTGGATCAATTTGAGATGATTTTGAAGTTAAAGTGTGGTGA 5142  
 DB 709 TCAATTTGATCATATCAAGTCAATTTTGTGATTTCTATCAAAACAAAGCTTAAAGTCA 768  
 QY 5143 AATATCAATTTCTTGGATTTGAATTTGTAACCGAATCGGTTATTTATTTATCTCA 5202  
 DB 769 ATTTGAATTTCTTTAGTATTTAGATTTGATTTGATTTGATTTGATTTGATTTGATTTG 828  
 QY 5203 AGAAAATTTCTCAAGAAATTTCTTAAGATTTCAAACTAGATGATC 5250  
 DB 829 AAGGAATTTGTCTAGACTTCTTCTGATTCAGAAATTTGAGGTTG 876

## RESULT 13

LOCUS CG821291/c 803 bp DNA linear GSS 18-NOV-2003  
 DEFINITION SOYBA90TF LargeInsertSoybeanGenLib Glycine max genomic clone B47J01.MT7011, genomic survey sequence.  
 ACCESSION CG821291  
 VERSION CG821291.1 GI:38279854  
 KEYWORDS  
 SOURCE  
 ORGANISM Glycine max (soybean)

Glycine max (soybean)  
 Glycine max

REFERENCE 1 (bases 1 to 803)  
 AUTHORS Shultz, J., Meksem, K., Shetty, J., Town, C.D., Koo, H., Potter, J., Wakefield, K., Zhang, H., Wu, C. and Lightfoot, D.A.  
 TITLE End sequencing of BACs comprising a provisional minimal tiling path from a fingerprint physical map of soybean (Glycine max) cultivar Forrest

JOURNAL  
COMMENT

Unpublished (2003)  
Other\_GSSs: SOTBA90TH  
Contact: Chris Town, J. L. Shultz and D. A. Lightfoot  
The Center of Excellence in Soybean Research, Teaching and Outreach, Southern Illinois University at Carbondale and Plant Genomics, The Institute for Genomic Research  
Room 176, Ag. Building, Mail Code 4415, Carbondale, IL 62901-4415, USA and 9712 Medical Center Drive, Rockville, MD 20850, USA  
Tel: 618 453 1797  
Email: jashultz@siu.edu, jshultz@siu.edu, cdtown@iigr.org (URL: http://bioinformatics.siu.edu)  
Clones approximating a minimum tiling path were re-arrayed from the library master plates prior to sequencing.  
For purposes of clone identification each clone name is a concatenation of the original clone location and its new location in the re-arrayed sequencing plates.  
Seq primer: GTATACGACTCCTATAGGGC  
Class: BAC ends.

FEATURES  
source

Location/Qualifiers  
1..803  
/organism="Glycine max"  
/mol\_type="genomic DNA"  
/cultivar="forrest"  
/db\_xref="taxon:3847"  
/clone="B47J01:MTF011"  
/note="LargeinsertSoybeanGenLib"  
/note="Organ: Leaves; Vector: pCD04541 (pBELOBAC11 EcoRI clones); Site: 1: BstXI; Soybean (Glycine max (L.) Merr.) cv. Forrest seeds were grown in greenhouse for fourteen days. Nuclei were isolated and embedded in agarose, restriction digested with Hind III BamHI or EcoRI, large size DNA fragments were ligated in vector V41 (pGLD04541) and electro transformed in DH10a cells. About 90,000 clones from BAC libraries were fingerprinted with HindIII and Hae III. Version 2 (automatic build) Contigs were built from 78,001 fingerprints. Contigs were manually examined to find the best non redundant tite path through the contigs representing about 13,000 clones. The clones were end sequenced."

## ORIGIN

Query Match 1.4%; Score 89.4; DB 29; Length 803;  
Best Local Similarity 45.9%; Pred. No. 7.8e-06;  
Matches 343; Conservative 0; Mismatches 401; Indels 3; Gaps 1;  
Db 4483 TAATGATCCGGATGATTAAGTCAATGCTGAATGAGAAATTTAGATCAAAAG 4542  
790 TGATCATCCGGAATGGGACAGCATGATTGATAAATACAGCTTTGGAGCATATAG 731  
Qy 4543 TGTTTAGGAAGAAATTCATTCGCCACCGGTGGAAGCCATATCTATGGGTGGGTACA 4602  
Db 730 TACTTGGAGCTAGTCTCTCTCCCGGAGAAAGAACTGTGGTGTGGTGGGCTA 671  
Qy 4603 TACTGAGAAATTTGATCTCTCAAGCGTGTTCGGAATCAAGTGTGTTCATG 4662  
Db 670 TGTATATAAGTTGGGCGCAATAGACAGATTGATAGCTCAAGGCTGATGTAAAC--- 614  
Qy 4663 CAACAGACAAAAGAAAATTTGATGACCCCTTTAGTGTATGTTCACTGTATAGA 4722  
Db 613 CAAAGGTTATCCGAGTTTATAGACCTAGACATAGAGATACCTTCTCCCTGTGGCTAA 554  
Qy 4723 TCTTGTGACATTAAGATTATGACATATAGAGTGTGGAATGAGATGACATTAACA 4782  
Db 553 AATTACTTCTATTTGAGCTTTCTTGGCATAGCTGCCATCCGTCATTTGGCACTATACA 494  
Qy 4783 TTTAGACGTGAGTGGCGATCTAAATGCCCTCTATTACTCATTCATCAATCAATTTATGT 4842  
Db 493 ATTGGAATTTAAATATGATCTCTACTTGAATAATTTAGAGAGAGATTATATAGAGA 434  
Qy 4843 CTTTCCCTCTAAATCAGACTTTGAAGAAAACCATGTTGGTATGTAAGAGCTTCTGT 4902  
Db 433 ACCACCGAATTTGTTGCTCAGGGGAGTCTAGCTTGTTCACAACTTGGAGGCTCT 374

Qy 4903 CTATGGTTAAACAGTCGGGTTTGAATGTATCAACATATCAAAAGATATTGAGAGA 4862  
Db 373 TTATGACTCAAAACATCCCGACAGAGCTTGTTGAAAATTCAGCTCAGTTGTACGC 314  
Qy 4963 CATTGGTTTACTCAAGTTTATACAAATATGTTTATTTTCACTTGAATTAAGAGG 5022  
Db 313 TTTTGGATGAAAGAAAGTGAAGACCATTCAGTTTTTATTTATGATACCTCATTTAG 254  
Qy 5023 ATCAGTAAATATTTAGTTTATATGTTATGATGATCTTATGTTGAGATTACAAAA 5082  
Db 253 TAGATGTGTTTACTTGTGTGTCTATGATGATGATGTTCTTACAGGAATATAGCAGA 134  
Qy 5083 AGTATGATATATTTTGTGATCATGATGATGATGATGATGATGATGATGATGATGAT 5142  
Db 193 AAAATGCTCAACTGAAGGAAACATTTATTAGCATTTTCACTAAGATGATGGGAAA 134  
Qy 5143 AATATCAATATCTGTTGTTGATGATGATGATGATGATGATGATGATGATGATGAT 5202  
Db 133 ACTTAATATTTTCTTGAATTTGAATTTGATGATGATGATGATGATGATGATGATGAT 74  
Qy 5203 AGAAAAATTTCTCAAGAAATTTACTTAA 5229  
Db 73 AAGAAAGTATCGTTAGACATCTTAA 47

RESULT 14  
CG943796 926 bp DNA linear GSS 15-DEC-2003  
LOCUS MBEMCS77F mbh2 Medicago truncatula genomic clone 84017, genomic  
DEFINITION survey sequence.  
ACCESSION CG943796  
KEYWORDS CG943796.1 GI:39848409  
SOURCE GSS.  
MEDICAGO truncatula (barrel medic)  
SOURCE Medicago truncatula  
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.  
1 (bases 1 to 926)  
TIGR  
Town, C.D., Shetty, V., Koo, H. and Feldblum, T.F.  
Sequencing of BAC ends from Medicago truncatula  
Unpublished (2003)  
Other\_GSSs: MBEMCS77R  
Contact: Chris Town

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523  
Fax: 301-838-0208  
Email: cdtown@iigr.org  
Seq primer: TGTAAACGACGCCAGT  
Class: BAC ends.  
Location/Qualifiers  
1..926  
/organism="Medicago truncatula"  
/mol\_type="genomic DNA"  
/cultivar="genotype A17"  
/db\_xref="taxon:3880"  
/clone="84417"  
/clone\_lib="mbh2"  
/note="Vector: pbeLOBAC11; Site\_1: HindIII; Site\_2: HindIII; Cook, D.R. and Kim, D.J., unpublished"

FEATURES  
source

Query Match 1.4%; Score 89.2; DB 29; Length 926;  
Best Local Similarity 47.9%; Pred. No. 8.2e-06;  
Matches 324; Conservative 0; Mismatches 343; Indels 9; Gaps 2;  
Db 4625 AAAGGTGTTGCGAAATCAAGTGTGTGTCATGCGCAACAGCAAAAGAAAATTG 4684  
Qy 4625 AAAGGTGTTGCGAAATCAAGTGTGTGTCATGCGCAACAGCAAAAGAAAATTG 4684  
Db 18 AAAGTTGCAAGAAACAAAGCAAGGTTAGTGTCTCAAGTTTAAACCAAGAAAGTATT 77  
Qy 4685 GATTATGACCTTTTATGTTAGTTACCTGTATATGATCTGTGATTAAGATTATG 4744

## ORIGIN



This Page Blank (uspto)